

*Πανεπιστήμιο Πελοποννήσου,
Σχολή Μηχανικών
Τμήμα Ηλεκτρολόγων Μηχανικών κ' Μηχανικών Υπολογιστών,*

Αλγόριθμοι Αναγνώρισης Προσώπων σε μη ελεγχόμενα Περιβάλλοντα

Ελισάβετ Αργυράκη,

Γιώργος Χαλδεάκης

Περιεχόμενα

Περίληψη	7
Μηχανική Όραση	8
Ορισμός.....	8
Ιστορική Αναδρομή	10
Εφαρμογές.....	11
Μηχανική Μάθηση	20
Ορισμός.....	20
Ιστορική Αναδρομή	23
Περιγραφή Αλγορίθμων Μηχανικής Μάθησης	24
Γραμμική Παλινδρόμηση (Linear Regression)	24
Λογιστική Παλινδρόμηση (Logistic Regression).....	27
Αποφάσεις Μέσω Δένδρου (Decision Tree)	28
Naive Bayes	33
k - Nearest Neighbors (kNN)	35
Support Vector Machines.....	36
Δείκτες Αξιολόγησης.....	40
Confusion Matrix	40
Sensitivity	42
Specificity	42
Τύποι I Λάθους	43
Τύπος II Λάθους.....	43
ROC/AUC.....	43
Ακρίβεια - Accuracy	44
Precision	44
Λοιπές Έννοιες Μηχανικής Μάθησης	45

Μη ισορροπημένο σετ δεδομένων - Imbalanced Dataset	45
Υπερπροσαρμογή - Overfitting	45
Μηχανική μάθηση vs βαθιά μάθηση	47
Τι είναι βαθιά μάθηση	47
Διάφορες μεταξύ μηχανικής και βαθιάς μάθησης	50
Νευρωνικά Δίκτυα.....	54
Εκπαίδευση ενός Νευρωνικού δικτύου	54
Εφαρμογή των ANN σε πραγματικό χρόνο	57
Πλεονεκτήματα Νευρωνικών Δικτύων	58
Κατηγορίες Νευρωνικών δικτύων	59
Συνελκτικά νευρωνικά δίκτυα	59
Πολυεπίπεδα Νευρωνικά δίκτυα Perceptron	60
Αναγνώριση προσώπου	76
Χαρακτηριστικά για αναγνώριση προσώπου	76
Ταξινομητές για αναγνώριση προσώπου	77
Τι είναι ταξινόμηση.....	77
Βασικές κατηγορίες ταξινομητών.....	80
Επίλογος	108
Βιβλιογραφία	Error! Bookmark not defined.
Βιβλιογραφία	109

Πίνακας Εικόνων

Εικόνα 1. ο συνδυασμός των πεδίων που αποτελείται η μηχανική όραση	11
Εικόνα 2. κατάθεση χειρόγραφης επιταγής σε ATM.....	12
Εικόνα 3. ανάλυση έκφρασης του πίνακα “Mona Lisa” του Leonardo Da Vinci	13
Εικόνα 4. εύρεση πληροφοριών των αποθηκευμένων φωτογραφιών στα γυαλιά της google	14
Εικόνα 5. Η στολή που φορούσε ο ηθοποιός Andy Serkis αποδύοντας τον smeagoi-Gollum στην ταινία lord of the rings.....	15
Εικόνα 6. το αυτοκίνητο της google	16
Εικόνα 7. η νοητή γραμμή που εμφανίζεται στην τηλεόραση μας κατά την προβολή του offside.....	17
Εικόνα 8. προβολή των κινήσεων μας στην οθόνη με τη βοήθεια του Microsoft Kinect	18
Εικόνα 9. προβολή τρισδιάστατων εικόνων του ανθρώπινου κρανίου.....	19
Εικόνα 10. πεδία έρευνας της μηχανικής μάθησης.....	22
Εικόνα 11. Απλή γραμμική παλινδρόμηση	25
Εικόνα 12. Διαφορές στις βέλτιστες γραμμές με και χωρίς outliers.....	26
Εικόνα 13. Σιγμοειδής Συνάρτηση.....	27
Εικόνα 14. Αναπαράσταση δένδρου αποφάσεων	29
Εικόνα 15. Σύστημα με υψηλό κέρδος πληροφορίας	30
Εικόνα 16. Σύστημα με χαμηλό κέρδος πληροφορίας.....	31
Εικόνα 17. Καθαρότητα ενός, υπό μελέτη, συστήματος.....	31
Εικόνα 18. Υπό συνθήκη πιθανότητα	34
Εικόνα 19 Πρόβλημα 5 κοντινότερων γειτόνων.....	35
Εικόνα 20. Γεωμετρική μορφή του υπερ επιπέδου.....	38
Εικόνα 21. Τιμή του Margin για τυχαίο υπερ επίπεδο.....	38
Εικόνα 22. Δείκτες – Metrics Ανάλογα το πρόβλημα που μελετάται.....	40
Εικόνα 23. Παράδειγμα Confusion Matrix.....	41
Εικόνα 24. False Positive και False Negative Περιπτώσεις	42
Εικόνα 25. ROC – AUC καμπύλες	43
Εικόνα 26. Accuracy - Prediction	44
Εικόνα 27. Overfitting	46

Εικόνα 28. Φαινόμενο Underfitting	47
Εικόνα 29. στα αριστερά παρατηρούμε ένα απλό νευρωνικό δίκτυο και δεξιά ένα νευρωνικό δίκτυο της βαθιάς μάθησης	50
Εικόνα 30. Η συσχέτιση μεταξύ της τεχνητής νοημοσύνης, της μηχανικής μάθησης και της βαθιάς μάθησης.....	51
Εικόνα 31. Η βασική διαφορά μεταξύ μηχανικής και βαθιάς μάθησης.....	52
Εικόνα 32. Διαφορές μεταξύ μηχανικής και βαθιάς μάθησης.....	53
Εικόνα 33. Τυπική δομή ενός Νευρωνικού Δικτύου	54
Εικόνα 34. Πολυεπίεδος Perceptron Εμπρόσθιας διάδοσης.....	60
Εικόνα 35. Η έννοια της Gradient Descent	64
Εικόνα 36. Κίνηση Gradient Descent σε τρισδιάστατο πρόβλημα.....	66
Εικόνα 37. Ολικό ελάχιστο – Βέλτιστη τιμή bias	67
Εικόνα 38. Ολικά/Τοπικά Ελάχιστα.....	68
Εικόνα 39. Η διαφορά μεταξύ ακραίων τιμών ρυθμού εκπαίδευσης.....	68
Εικόνα 40. Αναπαράσταση των διαφορετικών τύπων GD.....	69
Εικόνα 41. Τυχαίο FeedForward Νευρωνικό Δίκτυο	70
Εικόνα 42. Αρχικοποίηση βαρών και υπολογισμός των biases.....	71
Εικόνα 43. Αναπαράσταση του πρώτου βήματος της μετάδοσης του σφάλματος προς τα πίσω.....	73
Εικόνα 44. Μετάδοση του σφάλματος στο κρυφό επίπεδο	74
Εικόνα 45: Κανονική Κατανομή.....	81
Εικόνα 46: Δυσδιάστατο Πρόβλημα	82
Εικόνα 47: Διατήρηση της μίας διάστασης	82
Εικόνα 48: LDA Λύση	83
Εικόνα 49: Τροποποίηση του σετ δεδομένων μέσω της PCA.....	85
Εικόνα 50: Ιδιοδιανύσματα - Παράδειγμα γραμμικού μετασχηματισμού	88
Εικόνα 51: Ανεξάρτητες Πηγές Σήματος	90
Εικόνα 52: Γραμμικός Μετασχηματισμός των δύο πηγών	90
Εικόνα 53: Σχεδιασμός του box – Viola Jones	92
Εικόνα 54: Haar Χαρακτηριστικά.....	93
Εικόνα 55: Εντοπισμός Haar χαρακτηριστικών σε μία εικόνα	94
Εικόνα 56: Σύγκριση των τιμών της φωτεινότητας	95
Εικόνα 57: Κατασκευή της Integral Εικόνας.....	96

Εικόνα 58: Υπολογισμός της τιμής του feature του ROI.....	96
Εικόνα 59: Συντελεστές για τον κάθε weak classifier και η δημιουργία του ισχυρού	97
Εικόνα 60: Cascading διαδικασία	98
Εικόνα 61: Αποτελέσματα Εικόνας 1.....	99
Εικόνα 62: Αποτελέσματα Εικόνας 2.....	100
Εικόνα 63: Αποτελέσματα Εικόνας 3.....	101
Εικόνα 64: Rescaling της εικόνας.....	103
Εικόνα 65: P - Net.....	103
Εικόνα 66: R - Net.....	103
Εικόνα 67: O – Net.....	104
Εικόνα 68: Αποτελέσματα Εικόνας 1.....	105
Εικόνα 69: Αποτελέσματα Εικόνας 2.....	106
Εικόνα 70: Αποτελέσματα Εικόνας 3.....	106

Περίληψη

Στη συγκεκριμένη διπλωματική εργασία προχωρήσαμε στην μελέτη του τομέα της τεχνητής νοημοσύνης και ειδικότερα, την κατηγορία της μηχανικής όρασης (Computer Vision). Μετά την καταγραφή της απαραίτητης θεωρίας, προχωρήσαμε στην εκτέλεση ορισμένων πειραμάτων, με σημείο αναφοράς γνωστούς αλγορίθμους και είμασταν σε θέση να εντοπίσουμε τις βασικές διαφορές, τα πλεονεκτήματα και μειονεκτήματά τους.

Μηχανική Όραση

Ορισμός

Η μηχανική όραση είναι ένα επιστημονικό πεδίο που ανήκει στον κλάδο της τεχνητής νοημοσύνης. Στόχος της είναι να αναπαράγει αλγοριθμικά την αίσθηση της όρασης, συνήθως σε ηλεκτρονικό υπολογιστή ή ρομπότ. Σχετίζεται άμεσα με τη θεωρία και την τεχνολογία που εμπλέκονται στη σχεδίαση και την κατασκευή συστημάτων που λαμβάνουν και αναλύουν δεδομένα από ψηφιακές εικόνες.

Ως σημαντική επιδίωξη της μηχανικής όρασης είναι η εφαρμογή μοντέλων και θεωριών στην κατασκευή μηχανικών συστημάτων με τη δυνατότητα της όρασης.

Παραδείγματα εφαρμογών τέτοιων συστημάτων είναι:

1. Ο έλεγχος των διαδικασιών. (π.χ. ένα βιομηχανικό ρομπότ)
2. Η ανίχνευση των συμβάντων. (π.χ. η οπτική επιτήρηση ή παρακολούθηση πλήθους)
3. Η οργάνωση των πληροφοριών. (π.χ. η ευρετηριοποίηση βάσεων δεδομένων και ακολουθιών εικόνων)
4. Η εξομοίωση των αντικειμένων και των περιβαλλόντων. (π.χ. ανάλυση ιατρικών εικόνων ή η τοπογραφική εξομοίωση)
5. Η αλληλεπίδραση χρηστών με υπολογιστές. (π.χ. ως είσοδος σε μια συσκευή επικοινωνίας ανθρώπου / μηχανής)

Η μηχανική όραση περιγράφεται κυρίως ως συμπλήρωμα της βιολογικής όρασης, παρά ως αντίθετο της. Στη βιολογική όραση μελετώνται η οπτική αντίληψη στους ανθρώπους και τα ζώα με αποτέλεσμα να προκύπτουν μοντέλα για το πως αυτά λειτουργούν υπό φυσιολογικές διαδικασίες. Η μηχανική όραση μελετά και περιγράφει τεχνητά συστήματα όρασης που εφαρμόζονται σε λογισμικό ή σε υλικό υπολογιστών. Οπότε κρίνεται απαραίτητη η συνεργασία μεταξύ των δύο αυτών πεδίων για την μελλοντική τους ανάπτυξη.

Η βιολογική όραση που έχει ο κάθε άνθρωπος όμως δεν είναι ίδια με τη μέτρηση των ιδιοτήτων της εικόνας που επεξεργάζεται ο υπολογιστής. Ο ανθρώπινος εγκέφαλος αντιλαμβάνεται μια εικόνα τελείως διαφορετικά απ' ό,τι ένας υπολογιστής. Αυτό συμβαίνει κυρίως γιατί η βιολογική όραση είναι η κατασκευή μιας αντίληψης για το

τι υπάρχει στον έξω κόσμο, βασισμένη στις μετρήσεις που 'χει κάνει ο αισθητήρας εικόνας του ανθρώπινου εγκεφάλου. Συνήθως, αυτές οι μετρήσεις δημιουργούνται με την αντίληψη σκιών ή χρωμάτων τα οποία αντιλαμβάνεται διαφορετικά ο ανθρώπινος νους απ' ότι ένας ηλεκτρονικός υπολογιστής.

Αυτό που πραγματικά εννοείται όταν γίνεται αναφορά της μηχανικής όρασης, είναι έξυπνες κάμερες, οι οποίες έχουν υπολογιστική ικανότητα, να δίνουν την δυνατότητα σε μηχανές να αναγνωρίσουν το περιβάλλον στο οποίο βρίσκονται και να δράσουν αναλόγως σε όλες τις περιπτώσεις με άριστο ποσοστό επιτυχίας. Συνδυάζει τεχνικές όπως τον φωτισμό, την ηλεκτρονική, τον αυτοματισμό και την οπτική, συλλέγοντας πληροφορίες και καταλήγει σε αποφάσεις με βάση το ειδικό λογισμικό που προδιαθέτει. Είναι χρήσιμη σχεδόν σε όλες τις βιομηχανίες παραγωγής, διότι είναι βασική τεχνολογία στο βιομηχανικό αυτοματισμό αφού ελέγχει τις διαδικασίες παραγωγής μέσω της ερμηνείας των εικόνων που λαμβάνει.

Η μηχανική όραση χωρίζεται σε υποκατηγορίες για την καλύτερη κατανόηση και επίλυση των προβλημάτων της. Αυτές είναι:

- Η κατανόηση σκηνής
- Η ανίχνευση συμβάντων
- Η ανίχνευση κίνησης
- Η αναγνώριση αντικειμένων
- Η ευρετηριοποίηση
- Η αναγνώριση κίνησης
- Η τρισδιάστατη ανακατασκευή

Σε σχέση με άλλα τώρα, συγγενή γνωστικά πεδία, η μηχανική όραση διαφοροποιείται ως εξής:

- Η ψηφιακή επεξεργασία εικόνας (digital signal image) εξετάζει αλγορίθμους οι οποίοι δέχονται ως είσοδο εικόνες / βίντεο και παράγουν ως έξοδο εικόνες / βίντεο.

- Τα γραφικά υπολογιστή (computer graphics) εξετάζουν αλγορίθμους οι οποίοι δέχονται ως είσοδο συμβολικές περιγραφές οπτικών σκηνών και παράγουν ως έξοδο εικόνες / βίντεο (με ή χωρίς αλληλεπίδραση με τον χρήστη).
- Η μηχανική όραση εξετάζει αλγορίθμους οι οποίοι δέχονται ως είσοδο εικόνες / βίντεο και παράγουν συμβολικές περιγραφές των εν λόγω οπτικών σκηνών.
- Η μηχανική μάθηση εξετάζει αλγορίθμους οι οποίοι δέχονται ως είσοδο δεδομένα κάθε τύπου και τα ταξινομούν ή κατηγοριοποιούν σε ομάδες. Ένα σημαντικό μέρος της μηχανικής όρασης, είναι ουσιαστικά εφαρμογή της μηχανικής μάθησης σε ψηφιακά οπτικά δεδομένα.

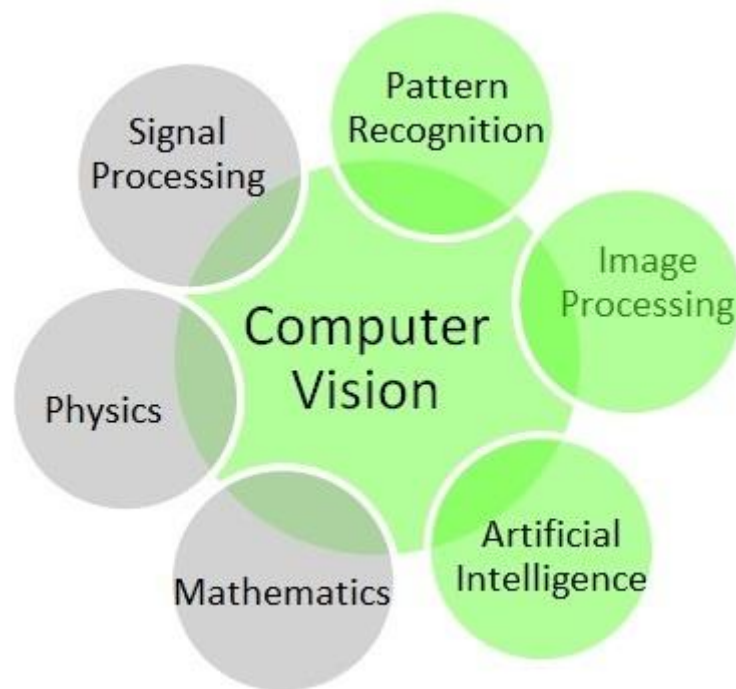
Ιστορική Αναδρομή

Η υπολογιστική ή μηχανική ή τεχνητή όραση (computer vision) είναι ένας σχετικά νέος κλάδος επιστήμης. Η ανάδειξή του έγινε μετά το 1980, που προέκυψε ως αποτέλεσμα επέκτασης του πεδίου της πληροφορικής, όπου καλείται ψηφιακή επεξεργασία εικόνας σε αλγορίθμους ανάλυσης και κατανόησης εικόνων. Πριν την περίοδο εκείνη, είχαν κατασκευαστεί μαθηματικά μοντέλα της φυσικής όρασης, κι είχαν γίνει οι πρώτες προσπάθειες για αναπαραγωγή της αίσθησης της όρασης σε αυτόνομα ρομπότ. Σχετιζόταν άμεσα την περίοδο εκείνη με τους όρους της ηλεκτρολογίας και της ρομποτικής λόγω της ραγδαίας βιομηχανικής εξέλιξης. Κατά τη διάρκεια της δεκαετίας του '80 και με την εμφάνιση της υπολογιστικής όρασης, αυτοί οι δύο όροι συγχωνεύτηκαν ως επιστημονικά πεδία, δημιουργώντας τον τομέα της τεχνητής νοημοσύνης. Αυτός ο τομέας είχε εφαρμογές όχι μόνο στην ρομποτική αλλά και σε δεκάδες ακόμα κλάδους.

Με το πέρασμα του χρόνου και φτάνοντας στη δεκαετία του '90, η μηχανική όραση είχε κάνει αισθητή την παρουσία της. Λόγω της μεγάλης και ξαφνικής ανάπτυξης της, συνδέθηκε με το γνωστικό πεδίο της μηχανικής μάθησης. Έδωσε σημαντικά αποτελέσματα, με αλγορίθμους όρασης πραγματικού χρόνου να υλοποιούνται ακόμα και σε φτηνά κινητά τηλέφωνα εξοπλισμένα με κάμερα. Έπαιξε τεράστιο ρόλο στην εξέλιξη της ενισχυμένης πραγματικότητας.

Μετά την εφεύρεση και παγκόσμια διάδοση των δυνατοτήτων που πρόσφερε η κάμερα της Microsoft, το Kinect, ένα καινοτόμο περιφερειακό διασύνδεσης μεταξύ

χρηστών και υπολογιστικών συστημάτων, αλλά και τη τρισδιάστατη προβολή οπτικού περιεχομένου, όπως η κινηματογραφική ταινία Avatar, η μηχανική όραση άρχισε να εξετάζει την αξιοποίηση εικόνων βάθους για την επίτευξη των στόχων της. Αυτό έγινε δυνατό με την χρήση στερεοσκοπικών καμερών και ξεχωριστών αισθητήρων.



Εικόνα 1. ο συνδυασμός των πεδίων που αποτελείται η μηχανική όραση

Εφαρμογές

Οι εικόνες και η απεικόνιση τους έχει διαδραστικό χαρακτήρα στην καθημερινότητα του ανθρώπου. Με την τεχνολογία που έχει αναπτυχθεί ραγδαία βρίσκονται παντού. Στις φωτογραφίες, στα βίντεο όπου και δίνεται η δυνατότητα διαμοίρασης τους στα κοινωνικά μέσα. Έτσι, κρίνεται αναγκαία η εύρεση εφαρμογών ώστε να γίνει εφικτή η επεξεργασία και η ανάλυση αυτού του τεράστιου όγκου δεδομένων. Για παράδειγμα, συστήματα που εξάγουν πληροφορίες από την απεικόνιση είναι η παρακολούθηση, η δημιουργία 3D(τριπλής διάστασης) παραστάσεων για ιατρικούς σκοπούς ή ακόμα και η απαθανάτιση κινουμένων εικόνων. Η μηχανική όραση βρίσκεται στα πιο απλά πράγματα του σήμερα που κάποτε θεωρούνταν δύσκολο να επιτευχθούν.

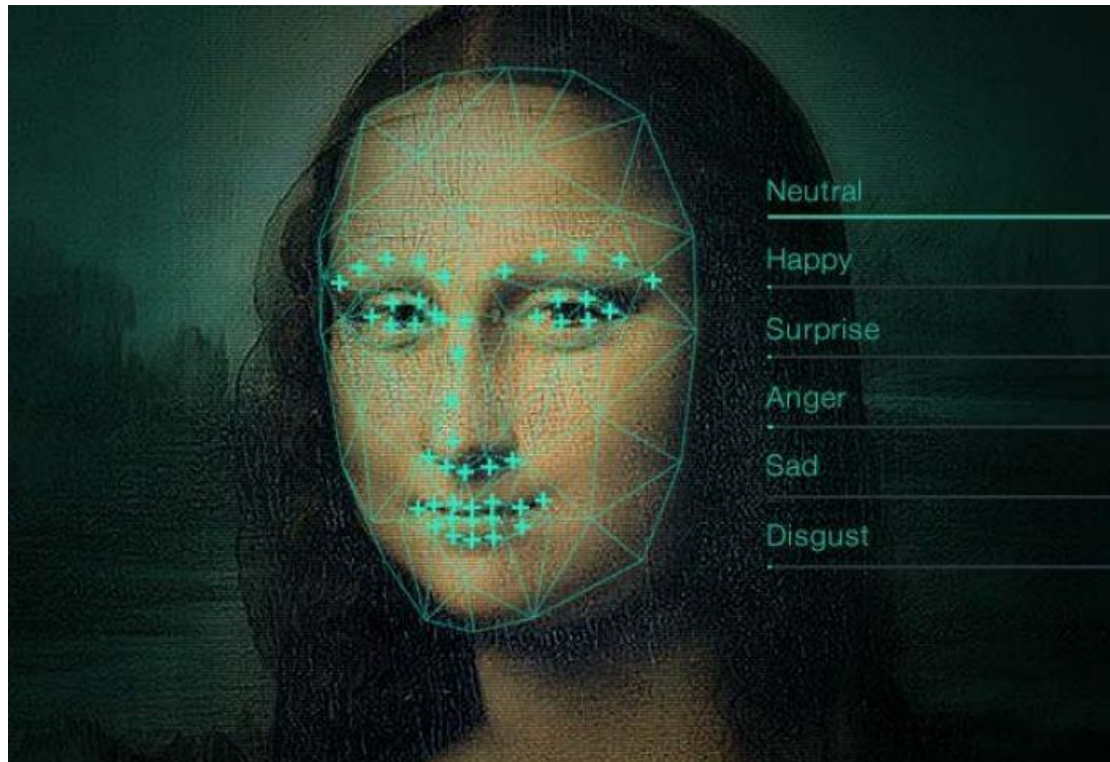
Διασπάται όμως κι αυτή σε πολλά διαφορετικά πεδία έρευνας. Ένα απ' αυτά είναι η οπτική αναγνώριση χαρακτήρων. Είναι η τεχνολογία που επιτρέπει την μετατροπή σκαναρισμένων αρχείων σε απλό κείμενο. Η οπτική αναγνώριση αρχείων, ή αλλιώς OCR (Optical Character Recognition), υπάρχει σε scanner που χρησιμοποιούμε όλοι στο σπίτι μας ή ακόμα και στο Adobe Acrobat το οποίο είναι ένα χρήσιμο εργαλείο στον υπολογιστή. Επίσης, βρίσκεται και σε πιο περίπλοκα συστήματα όπως στα μηχανήματα ανάληψης χρημάτων (ATM) , στα οποία υπάρχει η δυνατότητα κατάθεσης επιταγών με ποσό που έχει γραφεί δια χειρός, αλλά και στις κάμερες κυκλοφορίας όπου γίνεται αναγνώριση των πινακίδων κυκλοφορίας των οχημάτων. Μία πρόσφατη χρήση της οπτικής αναγνώρισης χαρακτήρων είναι η αναγνώριση του ταχυδρομικού κώδικα που ξεκίνησαν να χρησιμοποιούν οι ταχυδρομικές υπηρεσίες από χειρόγραφους φακέλους.



Εικόνα 2. κατάθεση χειρόγραφης επιταγής σε ATM

Μια ακόμη λειτουργία που ερευνάται στο πεδίο της μηχανικής όρασης είναι και η αναγνώριση προσώπου, όπου έχει γίνει απαραίτητη σχεδόν σε όλες τις κάμερες και τα βίντεο που απαθανατίζονται μέσω του smartphone κινητού. Οι περισσότερες ψηφιακές κάμερες μπορούν να εντοπίσουν πρόσωπα, αλλά η αλήθεια είναι πως μπορούν να κάνουν πολλά περισσότερα στις μέρες μας. Ένα απλό παράδειγμα είναι η

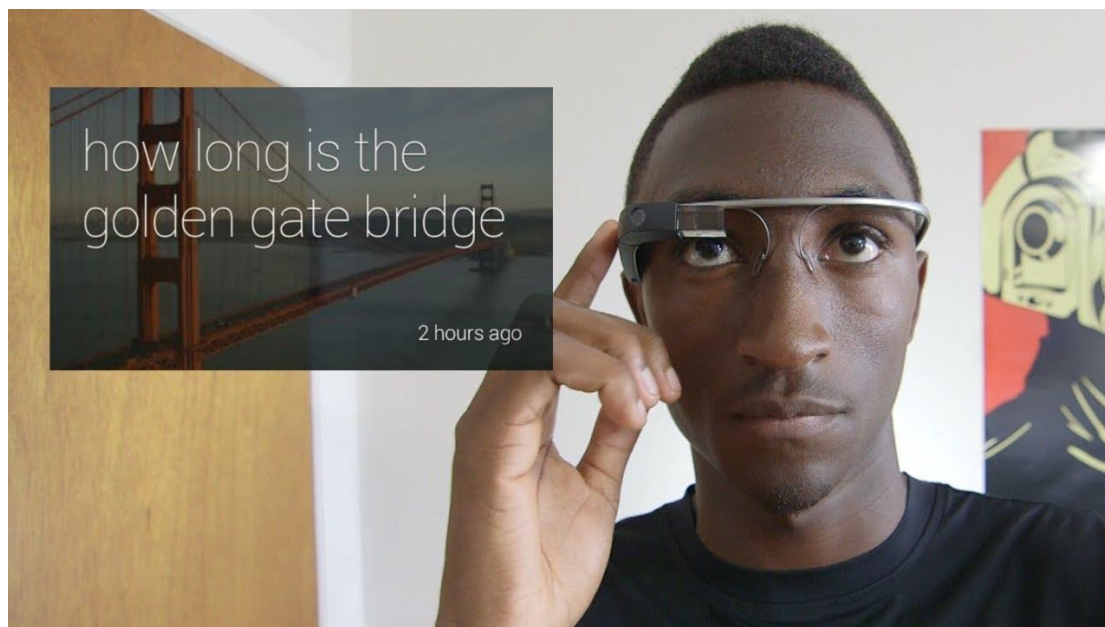
λήψη φωτογραφιών με την αναγνώριση χαμόγελου, όπου δίνει τη δυνατότητα στις φωτογραφικές μηχανές να τραβούν αυτόματα φωτογραφίες μόνο με την αναγνώριση χαμόγελου. Επίσης, μια ακόμη λειτουργία στις περισσότερες ψηφιακές κάμερες είναι η ειδοποίηση βλεφαρίσματος, αφού έχει αναγνωρισθεί πρώτα βλεφάρισμα κατά τη λήψη της φωτογραφίας. Επιπλέον, μπορούν να αναγνωρίσουν τον κάθε άνθρωπο ξεχωριστά, όπως στην σύνδεση χρήστη των windows με την απλή λήψη μιας φωτογραφίας του προσώπου του οποίου του ανήκει ο λογαριασμός.



Εικόνα 3. ανάλυση έκφρασης του πίνακα “Mona Lisa” του Leonardo Da Vinci

Εφάμιλλη λειτουργία αναγνώρισης με τις προηγούμενες είναι και η αναγνώριση αντικειμένων. Η εταιρεία Evolution Robotics έχει αναπτύξει ένα νέο τρόπο που αποσκοπεί στην προστασία εμπορικών προϊόντων από κλοπές στα εμπορικά καταστήματα. Το Lane Hawk, όπως και ονομάζεται, λειτουργεί με μία κάμερα στο κάτω μέρος του ταμείου η οποία αναγνωρίζει το προϊόν που περνά από μπροστά της. Με την τεχνολογία να εξελίσσεται ραγδαία η υπολογιστική ισχύς που χρειαζόταν για την αναγνώριση αντικειμένων έγινε μικρότερη και βρίσκεται ακόμη και στα κινητά τηλέφωνα. Δίνεται η δυνατότητα φωτογράφισης μνημείων ή αξιοθέατων και η αυτόματη λήψη πληροφοριών μέσω του διαδικτύου για το αντικείμενο που υπάρχει στη φωτογραφία. Πέραν όμως από τις λειτουργίες των smartphone, δίνεται η

δυνατότητα αναγνώρισης αντικειμένων και στα γυαλιά οράσεως. Τα γυαλιά της Google, ή αλλιώς Google Glass όπως αποκαλούνται, έχουν μία κάμερα πάνω τους η οποία είναι προσαρμοσμένη με την ορατότητα του ανθρώπινου ματιού και με την ίδια τεχνική αναγνώρισης αντικειμένων δίνει πληροφορίες για το συγκεκριμένο αντικείμενο.



Εικόνα 4. εύρεση πληροφοριών των αποθηκευμένων φωτογραφιών στα γυαλιά της google

Μία, όχι και τόσο γνωστή, χρήση της μηχανικής όρασης γίνεται και στα ειδικά εφέ που χρησιμοποιούνται συχνά στις ταινίες που προβάλλονται στον κινηματογράφο. Σκανάροντας τα χαρακτηριστικά ή την όψη ενός ανθρώπου κατασκευάζονται μοντέλα με απώτερο σκοπό την δημιουργία πολλαπλών ειδώλων του συγκεκριμένου προσώπου. Υπάρχει η δυνατότητα επεξεργασίας των συγκεκριμένων ειδώλων, όπως η διαφορετική φωτεινότητα από κάθε πλευρά ή κατεύθυνση, λόγω της κατασκευής αυτών των τρισδιάστατων μοντέλων (3D). Επίσης, η επιτυχής χρήση των CGI (Computer-Generated Imagery) εφέ είναι αποτέλεσμα επεξεργασίας της καταγραφής κίνησης που γίνεται μέσω ειδικών σημείων που τοποθετούνται πάνω στον ηθοποιό ή το αντικείμενο. Τα σημεία αυτά παρακολουθούνται συνεχώς από ειδικές κάμερες, οι οποίες βρίσκουν την τρισδιάστατη γεωμετρία, με αποτέλεσμα τα CGI να μην αποκόπτονται λόγω κάποιας κίνησης του ηθοποιού ή του αντικειμένου στο οποίο έχουν τοποθετηθεί. Τρισδιάστατα μοντέλα όμως δε χρησιμοποιούνται μόνο εκεί.

Κάνοντας λήψη μιας εναέριας εικόνας γίνεται εφικτή η αναπαράσταση τρισδιάστατων κτηρίων. Έτσι, η περιήγηση στο χώρο τριγύρω τους με τη βοήθεια εφαρμογών, όπως το Google Earth, γίνεται πλέον πραγματικότητα.



Εικόνα 5. Η στολή που φορούσε ο ηθοποιός Andy Serkis αποδύοντας τον smeagol-Gollum στην ταινία lord of the rings

Έχουν αναπτυχθεί τεχνολογίες που χρησιμοποιούν την μηχανική όραση στο αυτοκίνητο κάνοντας έτσι ευκολότερη και κυρίως ασφαλέστερη την εμπειρία οδηγώντας. Μπορούν να αναγνωρίσουν ακίνητα αντικείμενα όπως τα σήματα στο δρόμο αλλά και κινούμενα όπως τους πεζούς καθώς περνάνε μια διάβαση. Αναπτύσσονται επίσης και συστήματα που μειώνουν ταχύτητα και σταματάνε όταν αναγνωρίσουν ένα εμπόδιο. Το μείζων θέμα όμως που απασχολεί τον συγκεκριμένο κλάδο είναι η ανάπτυξη ενός μοντέλου αυτοκινήτου που έχει ήδη κατασκευαστεί, το οποίο συμπεριφέρεται αυτόνομα χωρίς τη βοήθεια οδηγού.

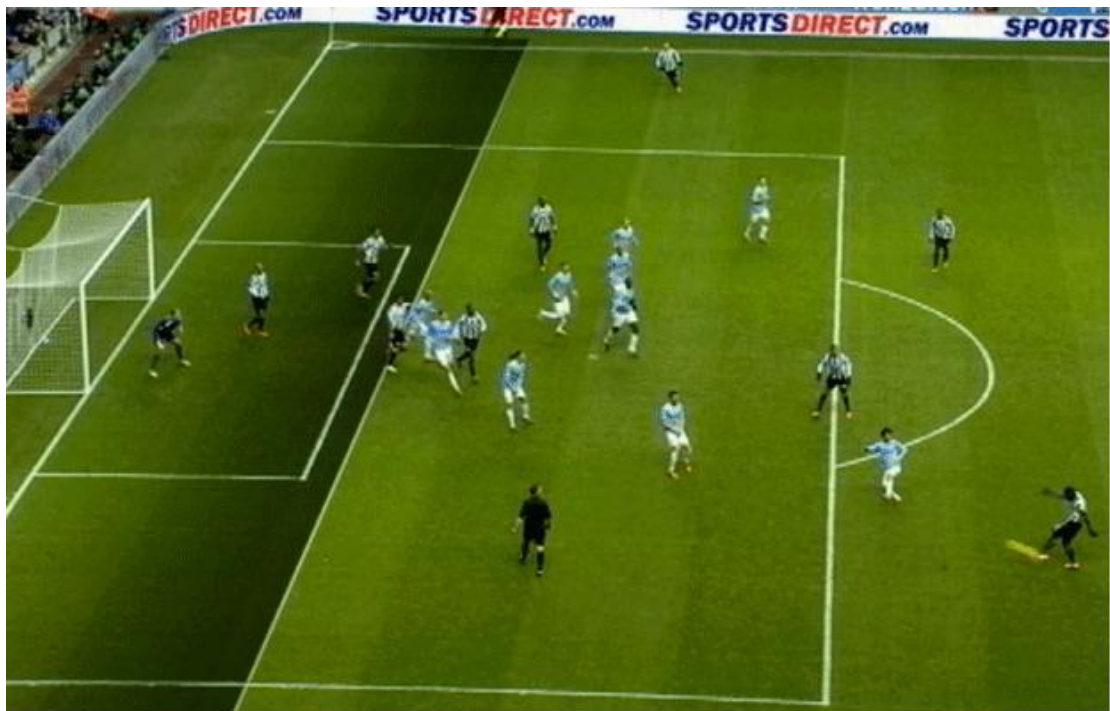
Κύριο, όμως, πρόβλημα στην ένταξη του στους δρόμους είναι η ανάληψη ευθυνών σε περίπτωση ατυχήματος με ό,τι οικονομικό ή πολιτικό κόστος συνεπάγεται.



Εικόνα 6. το αυτοκίνητο της google

Ένας χώρος δράσης αλλά και οικονομικής άνθισης της μηχανικής όρασης είναι και ο χώρος του πρωταθλητισμού και των σπορ. Οι τρόποι που χρησιμοποιείται η μηχανική όραση διακρίνονται ακόμη και στα πιο απλά πράγματα που γίνονται κατά τη διάρκεια ενός ποδοσφαιρικού αγώνα. Όπως για παράδειγμα, όταν ένας παίκτης μιας ομάδας υποπίπτει στο λεγόμενο “offside”, στην τηλεόραση εμφανίζεται μια νοητή γραμμή που σχηματίζεται στον τελευταίο αμυντικό της αντίπαλης ομάδας, την οποία υποτίθεται δεν έπρεπε να περάσει. Αυτή η γραμμή υπολογίζεται έτσι ώστε να μην διαπερνά τον αμυντικό της ομάδας, όπως επίσης να τον ξεχωρίζει από το υπόλοιπο γήπεδο αλλά και το φόντο στην εικόνα που μεταδίδεται στην τηλεόραση.

Επίσης, η μηχανική όραση μπορεί να αποβεί πολύ χρήσιμη και στους αγώνες τένις, όπου ειδικές κάμερες που αναγνωρίζουν το μπαλάκι μπορούν να διαπιστώσουν αν χτύπησε τελευταία φορά μέσα ή έξω απ’ τα όρια του γηπέδου αλλά και να υπολογίσουν την ταχύτητα του.



Εικόνα 7. η νοητή γραμμή που εμφανίζεται στην τηλεόραση μας κατά την προβολή του offside

Κάτι ακόμα που ‘χει αλλάξει αισθητά η μηχανική όραση είναι την εμπειρία στα ηλεκτρονικά παιχνίδια. Η Nintendo, μία από τις πιο κερδοφόρες εταιρείες πώλησης ηλεκτρονικών παιχνιδιών, λάνσαρε μια κονσόλα στην αγορά η οποία άλλαξε τον τρόπο ψυχαγωγίας, το Nintendo Wii. Η συγκεκριμένη κονσόλα περιέχει ένα τηλεχειριστήριο, το οποίο έχει ενσωματωμένο σύστημα κάμερας. Το σύστημα αυτό εντοπίζεται από έναν αισθητήρα κίνησης που βρίσκεται στην κονσόλα και ανάλογα την κίνηση που γίνεται με το χέρι είναι ορατό το αποτέλεσμα της στην οθόνη. Έτσι, δημιουργήθηκε μια πιο διαδραστική μέθοδος διασκέδασης που επέτρεπε στο χρήστη να γίνει πιο ενεργητικός.

Η πραγματική ανατροπή, που έγινε σ’ αυτό το είδος ψυχαγωγίας αλλά και σε υπόλοιπες διαδραστικές εφευρέσεις στην ρομποτική, είναι το Microsoft Kinect. Το Microsoft Kinect είναι ένας αισθητήρας βάθους που δημιουργήθηκε για την κονσόλα του Xbox. Δημιουργεί ασπρόμαυρες εικόνες που αντικατοπτρίζουν την απόσταση μεταξύ του χρήστη και των υπόλοιπων αντικειμένων στο χώρο. Όσο πιο σκούρο είναι το χρώμα τόσο πιο μακριά βρίσκεται το αντικείμενο, ενώ αντίθετα όσο πιο φωτεινό είναι τόσο πιο κοντά είναι. Το πιο σημαντικό όμως που κάνει το Microsoft Kinect είναι η δυνατότητα παραγωγής σκελετικών περιγραφών. Χρησιμοποιεί έναν

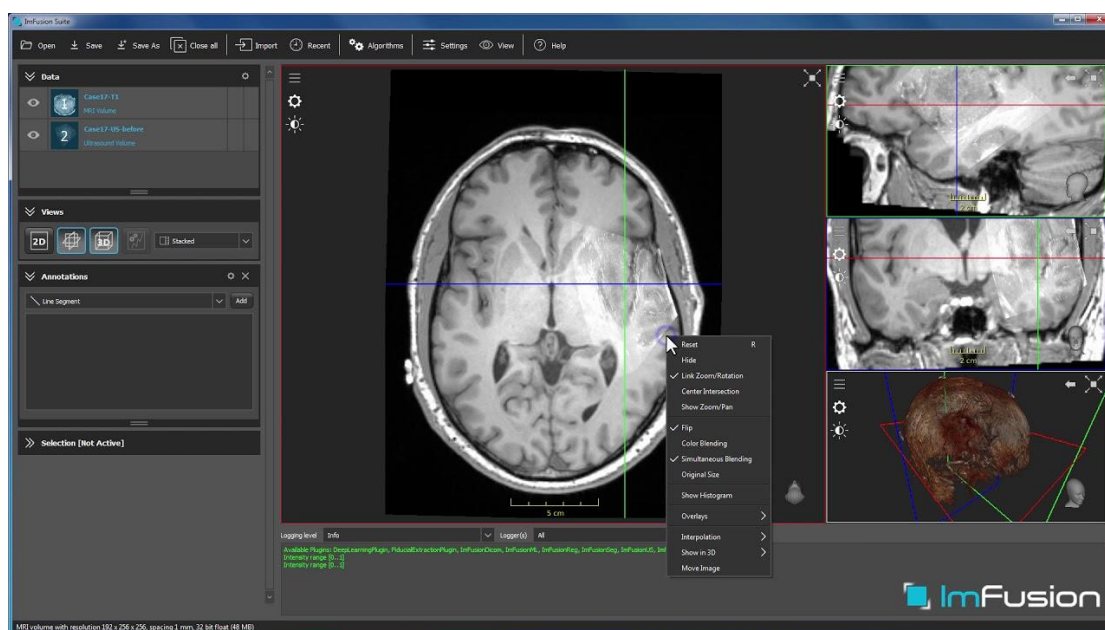
συνδυασμό τεχνικών της μηχανικής μάθησης (machine learning) βασισμένους στην μηχανική όραση. Ανακτώντας, λοιπόν, τη γεωμετρία αυτών των σκελετικών περιγραφών από ανθρώπους που βρίσκονται στην εικόνα βάθους, καταγράφουν κάθε κίνηση τους στιγμιαία στην οθόνη της τηλεόρασης. Έτσι, καθιστά άκρως ευχάριστη και ψυχαγωγική την διαδραστικότητα των εφαρμογών και των παιχνιδιών του που προσφέρει. Μεγάλη συμβολή όμως είχε και στην ρομποτική. Μέσω του Microsoft Kinect, πολλά ρομπότ έμαθαν να συλλέγουν τις πληροφορίες από το βάθος του περιβάλλοντος και την σκελετική περιγραφή των ανθρώπων. Έπειτα να τις ερμηνεύουν με βάση την ανθρώπινη καθημερινότητα και να δρουν αναλόγως, όπως για παράδειγμα έναν χαιρετισμό ή μια αγκαλιά. Ουσιαστικά, το Microsoft Kinect μπορεί δημιουργήθηκε για την ψυχαγωγία και την εξέλιξη των ηλεκτρονικών παιχνιδιών, αλλά είναι ένα αναγκαίο και χρήσιμο εργαλείο για πολλά περισσότερα πράγματα.



Εικόνα 8. προβολή των κινήσεων μας στην οθόνη με τη βοήθεια του Microsoft Kinect

Η μηχανική όραση έχει συμβάλει επίσης στην ασφάλεια και την υγεία των πολιτών. Σε πολυσύχναστα μέρη, όπως αεροδρόμια και εμπορικά κέντρα, έχουν τοποθετηθεί κάμερες ασφαλείας, όπου εντοπίζουν οχήματα και ανθρώπους που δε θα έπρεπε να είναι σε συγκεκριμένα σημεία. Μία πιο άμεση επίδραση της μηχανικής όρασης είναι

στην υγεία. Η λήψη τρισδιάστατων εικόνων (3D Imaging), η μαγνητική τομογραφία (MRI) και η αξονική τομογραφία (CT scan) είναι τεχνικές που έχουν αναπτυχθεί με τη βοήθεια της μηχανικής όρασης κι έχουν βοηθήσει στην αντιμετώπιση ιατρικών προβλημάτων. Η συμβολή τους στον τομέα της υγείας είναι μεγάλη, λόγω της γρήγορης διάγνωσης και απεικόνισης του ιατρικού προβλήματος που πάσχει ο ασθενής. Επίσης στον ειδικό ιατρικό τομέα της μικροχειρουργικής, 3D εικόνες βοηθούν στην αναλυτική και ακριβή απεικόνιση των λεπτομερειών του σημείου το οποίο υπόκειται σε χειρουργείο. Έτσι, η δουλειά του χειρουργού γίνεται ευκολότερη και πιο αποτελεσματική.



Εικόνα 9. προβολή τρισδιάστατων εικόνων του ανθρώπινου κρανίου

Τα περισσότερα επιτεύγματα της μηχανικής όρασης που έχουν συμβάλει σε όλους αυτούς τους τομείς είναι νέα στην ιστορία της τεχνολογίας. Αυτό σημαίνει πως το πεδίο της μηχανικής όρασης αλλάζει διαρκώς, με αποτέλεσμα αυτό να ‘χει αντίκτυπο στην καθημερινότητα του ανθρώπου. Εφευρίσκοντας νέες εφαρμογές που αλλάζουν το μέλλον, ανοίγουν νέες θέσεις εργασίας που ωθούν στην αύξηση της επιχειρηματικότητας. Κρίνεται αναγκαία, λοιπόν, η έμπνευση, η βοήθεια και η ώθηση νέων ανθρώπων σ’ αυτόν το καινούριο κλάδο τεχνολογίας, ώστε να βοηθήσουν στην εξέλιξη της ανθρωπότητας στο μέλλον.

Μηχανική Μάθηση

Η συνεχής ανάπτυξη της τεχνολογίας έχει φέρει αλλαγές στις ζωές των ανθρώπων σε όλα τα επίπεδα, έχει απασχολήσει ιδιαίτερα επιστήμονες και ερευνητές και έχουν προκύψει σημαντικές ανακαλύψεις με στόχο τόσο την βελτίωση της ανθρώπινης ύπαρξης, όσο και την κατανόηση των υπολογιστικών συστημάτων.

Κάνοντας λόγο για κατανόηση των υπολογιστικών συστημάτων, δεν θα μπορούσαμε να μην αναφέρουμε τον όρο της τεχνητής νοημοσύνης, η οποία αναφέρεται στον κλάδο της πληροφορικής που ασχολείται με την σχεδίαση και την υλοποίηση υπολογιστικών συστημάτων (μηχανήματα) που μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς τα οποία υπονοούν έστω και στοιχειώδη ευφυΐα.

Συνεπώς, η ανάγκη για την ανθρώπινη συμπεριφορά των υπολογιστικών συστημάτων προέκυψε σαν επιστημονικό εγχείρημα η μηχανική μάθηση που καλείται να καλύψει το χάσμα που δημιουργεί η τεχνητή νοημοσύνη με τα υπολογιστικά συστήματα.

Ορισμός

Μηχανική μάθηση ονομάζεται ο κλάδος της τεχνητής νοημοσύνης που ασχολείται με την μελέτη ή την κατασκευή συστημάτων που έχουν την ικανότητα να «μαθαίνουν» από ένα πλήθος δεδομένων. Συγκεκριμένα, η μηχανική μάθηση στην προσπάθειά της να κάνει ένα σύστημα να «μάθει» από τα δεδομένα, κατασκευάζει αλγορίθμους που καλούνται να εξυπηρετήσουν τον σκοπό αυτό.

Οι αλγόριθμοι αυτοί έχουν ως λειτουργία την κατασκευή μοντέλων από πειραματικά δεδομένα, προσπαθώντας να κάνουν μία πρόβλεψη που βασίζεται στα δεδομένα που τους δίνουν ή να βγάλουν κάποιο αποτέλεσμα από αυτά.

Δεν θα μπορούσαμε να μην αναφέρουμε τους ορισμούς ερευνητών, αναφερόμενοι στην μηχανική μάθηση, κάνοντας λόγο για τους:

- Arthur Samuel (1987), που ορίζει την μηχανική μάθηση ως: «το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί»,
- Tom M. Mitchell (1997), που έδωσε για την μηχανική μάθηση έναν ορισμό που χρησιμοποιείται ευθέως: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μία κλάση εργασιών T και ένα μέτρο

επίδοσης P, αν η επίδοσή του σε εργασίες της κλάσης Τα, όπως αποτιμάται από το μέτρο P, βελτιώνεται με την εμπειρία E»,

- Τέλος, οι Ian H. Witten και Eide Frand (2000) με την σειρά τους δίνουν το δικό τους ορισμό της μηχανικής μάθησης: «Ένα σύστημα μαθαίνει όταν αλλάζει τη συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον».

Η μηχανική μάθηση ακόμη, συνδέεται στενά με τον κλάδο της υπολογιστική στατιστικής. Ενώ η μηχανική μάθηση όπως αναφέραμε παραπάνω, στηρίζεται σε αλγορίθμους που μπορούν να μάθουν από δεδομένα χωρίς να έχουν προγραμματιστεί ρητά, η υπολογιστική στατιστική έρχεται να επικεντρωθεί στην πρόβλεψη μέσω της χρήσης υπολογιστών αλλά και στη σχέση των μεταβλητών στα δεδομένα με την μορφή μαθηματικών εξισώσεων.

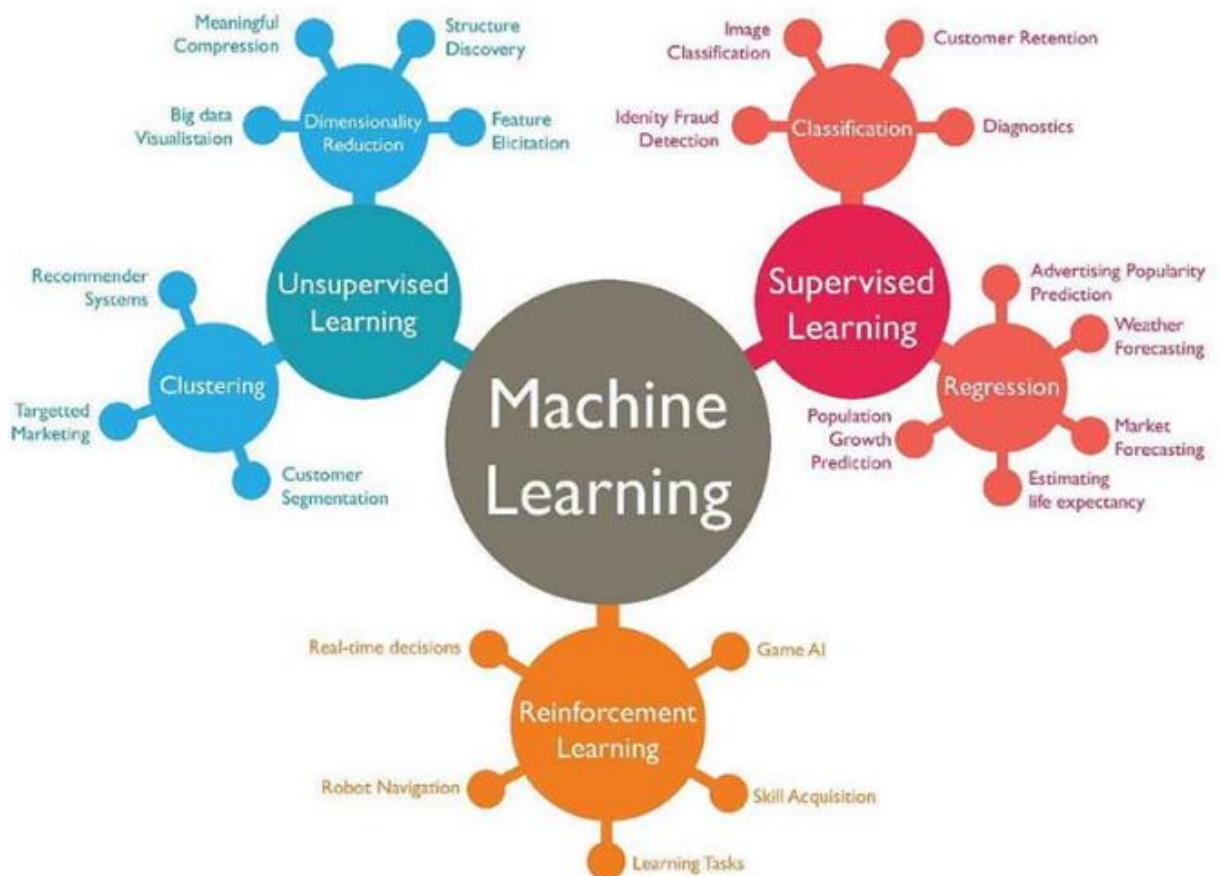
Η μηχανική μάθηση έχει αναπτύξει τρεις κατηγορίες ανάλογα με τον τρόπο που μελετά μια κατάσταση:

1. Μάθηση με επιτήρηση (Supervised Learning): Οι αλγόριθμοι μάθησης στέλνουν ως εισόδους δεδομένα και παράγουν αποτελέσματα, οπότε στόχους του υπολογιστικού συστήματος είναι να αντιστοιχεί την κάθε είσοδο με το αντίστοιχο αποτέλεσμα.
2. Μάθηση χωρίς επιτήρηση (Unsupervised learning): Οι αλγόριθμοι μάθησης πρέπει να βρουν την δομή των δεδομένων, χωρίς να γνωρίζουν το αποτέλεσμα.
3. Μάθηση με ενίσχυση (Reinforcement Learning): Οι αλγόριθμοι μάθησης αλληλεπιδρούν άμεσα με το περιβάλλον και τα δεδομένα που παίρνουν με σκοπό να φτάσουν στο επιθυμητό αποτέλεσμα.

Επιπλέον, ανάμεσα στη μάθηση με επιτήρηση και στη μάθηση χωρίς επιτήρηση, προκύπτει και η κατηγορία της ημι-επιτήρησης μάθησης, όπου οι αλγόριθμοι μάθησης δίνουν ανεπαρκείς εισόδους δεδομένων. Άρα κάποιοι από τους στόχους των αποτελεσμάτων λείπουν. Η μεταγωγή είναι ένα μέρος της συγκεκριμένης κατηγορίας που δίνει έμφαση στο γεγονός ότι ένα μέρος των στόχων λείπουν.

Τα προβλήματα της μηχανικής μάθησης μπορούν να προκύψουν αν θεωρήσουμε το επιθυμητό αποτέλεσμα του συστήματος της μηχανικής μάθησης:

1. Ταξινόμηση (Classification): Τα δεδομένα που έρχονται ως είσοδοι χωρίζονται σε δύο ή περισσότερες κλάσεις, άρα η μηχανή πρέπει να αντιστοιχίσει σε μία οι περισσότερες κλάσεις.
2. Παλινδρόμηση (Regression): Τα αποτελέσματα που παίρνουμε είναι συνεχή αλλά όχι αντιληπτά.
3. Συσταδοποίηση (Clustering): Τα δεδομένα εισόδων πρέπει να χωριστούν σε ομάδες που δεν είναι γνωστές.
4. Εκτίμηση πυκνότητας πιθαν. (Probability Density Estimation): Πρέπει να βρεθεί χώρος για τα δεδομένα εισόδου.
5. Το πρόβλημα της μείωσης διαστασιμότητας: Τα δεδομένα που εισέρχονται απλοποιούνται ώστε να αντιστοιχούν σε έναν μικρότερο χώρο.



Εικόνα 10. πεδία έρευνας της μηχανικής μάθησης

Ιστορική Αναδρομή

Το ζήτημα της κατασκευής μηχανών που ήταν σε θέση να μαθαίνουν από τα δεδομένα που τους έστελναν, ήταν το κυρίως ζήτημα που απασχολούσε τους ερευνητές και ήταν κάτι που προσπάθησαν να επιλύσουν με συμβολικές μεθόδους και με τα λεγόμενα νευρωνικά δίκτυα. Στην ουσία ήταν δίκτυα που αποτελούνταν από αλγόριθμους που μπορούσαν να επιλύσουν υπολογιστικά προβλήματα. Οι αλγόριθμοι αυτοί ήταν βασισμένοι στην στατιστική, κάτι που δημιουργούσε πρόβλημα στην φύση των υπολογιστών, που ήταν προγραμματισμένοι να επιλύουν λογικές πράξεις και να κατανοούν το δυαδικό σύστημα.

Το 1980, έμπειρα συστήματα έκαναν την εμφάνισή τους στον χώρο της τεχνητής νοημοσύνης, οπότε η στατιστική πλέον δεν αποτελούσε πρόβλημα στην συμβολική αναπαράσταση προβλημάτων. Ο ρόλος της θα αφορούσε από δω και πέρα μόνο την ανάκτηση πληροφοριών. Ακόμα τα νευρωνικά δίκτυα που αναφέραμε παραπάνω, εφόσον ήταν μέρος της στατιστικής στην επίλυση των προβλημάτων, σταμάτησαν να χρησιμοποιούνται.

Φτάνοντας στο 1990 και μετά από όλη αυτή την ανασυγκρότηση, η μηχανική μάθηση απέκτησε ενεργό ρόλο στην επίλυση προβλημάτων πρακτικής φύσης, με τον συμβολικό χαρακτήρα που της παρείχε η τεχνητή νοημοσύνη. Δόθηκε επίσης έμφαση στις μεθόδους της στατιστικής και της θεωρίας των πιθανοτήτων.

Η μηχανική μάθηση και η εξόρυξη δεδομένων συχνά χρησιμοποιούν τις ίδιες μεθόδους, ωστόσο έχουν σημαντικές διαφορές.

Εξετάζοντάς τα ξεχωριστά μπορούμε να πούμε πως η μηχανική μάθηση επικεντρώνεται στην πρόβλεψη πάνω σε γνωστές πληροφορίες που αποδίδονται στο σύνολο της εκπαίδευσης. Από την άλλη, η εξόρυξη δεδομένων επικεντρώνεται στο να βρεθούν πληροφορίες τις οποίες δεν γνωρίζουμε εκ των υστέρων.

Η επικάλυψη αυτών των δυο, έχει να κάνει ακριβώς με ότι αναφέρθηκε παραπάνω. Η μηχανική μάθηση αξιολογείται ως την ικανότητα αναπαραγωγής γνώσης που ήδη γνωρίζουμε, ενώ αντίστοιχα η ανακάλυψη αυτής της γνώσης, ή αλλιώς η εξόρυξη δεδομένων, στοχεύει στο να ανακαλυφθεί η γνώση που δεν γνωρίζουμε από την αρχή. Για τον λόγο αυτό μία μέθοδος επιτηρούμενης μάθησης μπορεί να φέρει

αποτελέσματα, ενώ σε μία τυπική διεργασία εξόρυξης δεδομένων η συγκεκριμένη μέθοδος δεν μπορεί να φέρει αποτελέσματα, λόγω της μη διαθεσιμότητας συνόλου εκπαίδευσης.

Ακόμη, η μηχανική μάθηση συνδέεται με την βελτιστοποίηση. Πολλά προβλήματα μάθησης διατυπώνονται ως η ελαχιστοποίηση της συνάρτησης απώλειας από ένα σύνολο δεδομένων. Η συνάρτηση απώλειας αναπαριστά την διαφορά που έχει η πραγματική κατάσταση ενός προβλήματος, με την κατάσταση του μοντέλου εκπαίδευσης που έχουμε διαλέξει. Η διαφορά τους έχει να κάνει με το γεγονός πως ενώ οι αλγόριθμοι βελτιστοποίησης μπορούν να ελαχιστοποιήσουν την απώλεια ενός συνόλου εκπαίδευσης, η μηχανική μάθηση στοχεύει στην ελαχιστοποίηση απώλειας σε άγνωστες καταστάσεις.

Τέλος, η μηχανική μάθηση συνδέεται και με την στατιστική. Δεν θα μπορούσαμε να μην αναφέρουμε τους:

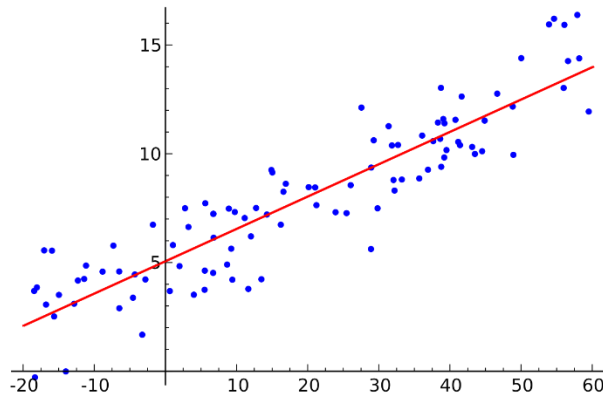
- Michael Jordan, ο οποίος αναφέρει πως οι ιδέες της μηχανικής μάθησης προϋπάρχουν στην στατιστική, και ο
- Leo Breiman, που επισημαίνει δυο υποδείγματα στατιστικής μοντελοποίησης: το μοντέλο δεδομένων και το αλγοριθμικό μοντέλο.

Περιγραφή Αλγορίθμων Μηχανικής Μάθησης

Ένα αρκετά σημαντικό κομμάτι, που δεν πρέπει να παραληφθεί από την ανάλυσή μας, αποτελεί η περιγραφή των βασικών αλγορίθμων που χρησιμοποιούνται για τα προβλήματα της μηχανικής μάθησης (machine learning), των βασικών εννοιών που τους διέπουν καθώς και μίας συνοπτικής αναφοράς, εκάστου, μαθηματικού υποβάθρου.

Γραμμική Παλινδρόμηση (Linear Regression)

Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται με σκοπό την εύρεση γραμμικής σχέσης μεταξύ των μεταβλητών που επιδιώκουμε να προβλέψουμε και των εισόδων τους. Η ίδια μπορεί είτε να αφορά μία έξοδο ή πολλαπλές. Κατά την απλή, (Simple Linear Regression, 2017), επιδιώκεται η εύρεση σχέσης μεταξύ δύο συνεχών μεταβλητών (μίας εξόδου και μίας εισόδου)



Εικόνα 11. Απλή γραμμική παλινδρόμηση

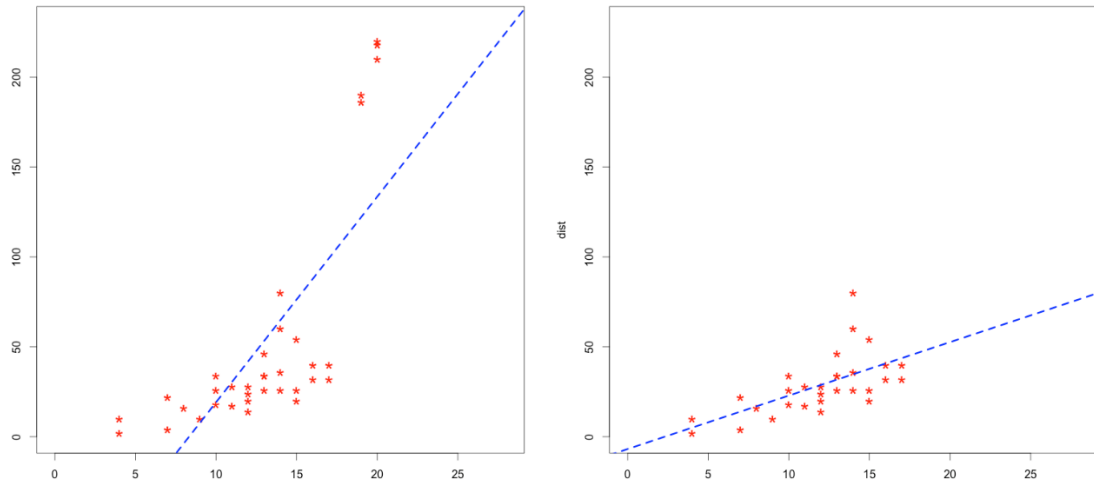
Η εξίσωση που περιγράφει την γραμμή που φαίνεται στην εικόνα 11, έχει την παρακάτω μορφή:

$$Y = a + bX$$

με X την ανεξάρτητη και Y την εξαρτημένη μεταβλητή

Το σημαντικότερο ζήτημα εντοπίζεται στο πώς μπορεί να ταιριάξει (fit) κάποιος αυτή την γραμμή στο σύνολο των σημείων μίας δυοδιάστατης απεικόνισης. Ο πιο γνωστός αλγόριθμος αφορά την μέθοδο των ελαχίστων τετραγώνων (Least Squares), κατά τον οποίο η βέλτιστη γραμμή είναι εκείνη που ελαχιστοποιεί το άθροισμα των τετραγώνων των αποστάσεων κάθε σημείου από την γραμμή (χρησιμοποιούμε το τετράγωνο ώστε να απαλείψουμε την επιρροή του πρόσημου στις πράξεις μας). Υπάρχουν και άλλοι μέθοδοι για την βελτιστοποίηση των παραμέτρων όπως η Gradient Descent και το Regularization οι οποίες, όμως, θα αναλυθούν παρακάτω.

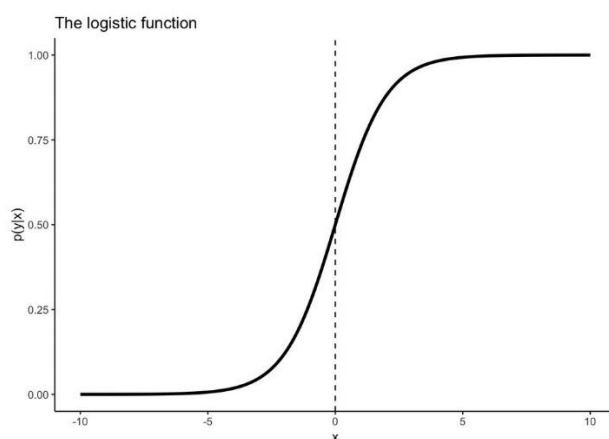
Πριν ολοκληρώσουμε την αναφορά στην συγκεκριμένη μέθοδο πρέπει να αναφέρουμε ότι παρόλο που αποτελεί μία απλή και οικονομική, από θέμα υπολογιστικού κόστους προσέγγιση, η οποία όμως έρχεται με τα αρνητικά της. Συγκεκριμένα, είναι δυνατόν, outlier τιμές να επηρεάσουν σημαντικά τα αποτελέσματα, δημιουργώντας σύγχυση. Για τον λόγο αυτό απαιτείται προσεκτική προ επεξεργασία των δεδομένων πριν λάβει χώρα οποιαδήποτε σοβαρή προσπάθεια κατηγοριοποίησης.



Εικόνα 12. Διαφορές στις βέλτιστες γραμμές με και χωρίς outliers

Λογιστική Παλινδρόμηση (Logistic Regression)

Ο συγκεκριμένος αλγόριθμος (Hoffman, 2015), αφορά την κατηγορία των προβλημάτων παλινδρόμησης κατά τα οποία η εξαρτημένη μεταβλητή (αυτή, δηλαδή που επιδιώκουμε να προβλέψουμε) μπορεί να λάβει δυαδικές τιμές (0 ή 1, όχι/ναι κτλ). Η ονομασία του, οφείλεται στην συνάρτηση γύρω από την οποία εκτυλίσσεται η συγκεκριμένη ανάλυση, δηλαδή την λογιστική (logistic) συνάρτηση, η οποία αλλιώς ονομάζεται και σιγμοειδής, λόγω του σχήματος της:



Εικόνα 13. Σιγμοειδής Συνάρτηση

Ο μαθηματικός τύπος της είναι ο παρακάτω, κατά τον οποίο, για οποιαδήποτε πραγματικό αριθμό, υπάρχει μία αντίστοιχη μεταξύ του 0 και 1.

$$\frac{1}{1 + e^{-x}}$$

Η συγκεκριμένη μέθοδος παραμένει μία γραμμική ανάλυση, της οποίας όμως η προβλέψεις, μετασχηματίζονται μεταξύ 0 και 1 μέσω της παραπάνω συνάρτησης. Λαμβάνοντας υπόψιν την εξίσωση της γραμμικής παλινδρόμησης, η παραπάνω θα μπορούσε να γραφτεί και ως:

$$\frac{e^{a+bX}}{e^{a+bX} + 1}$$

Στην συγκεκριμένη μέθοδο, πρέπει να κρατήσουμε στο μυαλό μας, ότι αναφερόμαστε σε μοντελοποίηση της πιθανότητας ενός ενδεχομένου (να βρίσκεται πχ σε μία κατάσταση A) που μπορούμε να την ονομάσουμε P(x).

Αν ήταν δηλαδή να την αποτυπώσουμε θα ίσχυε η παρακάτω έκφραση:

$$P(x) = \frac{e^{a+bX}}{e^{a+bX} + 1}$$

Γνωρίζοντας ότι $Y=a+bX$ θα μπορούσαμε το παραπάνω να το γράψουμε και ως:

$$\ln\left(\frac{P(x)}{1 - P(x)}\right) = a + bX = Y$$

Δικαιολογώντας τον ισχυρισμό ότι η παραπάνω πράγματι είναι γραμμική ανάλυση, όπου εδώ όμως λαμβάνεται υπόψιν ο λογάριθμος του ενδεχόμενου ύπαρξης αυτής της κατάστασης.

Για τον υπολογισμό των τιμών του b , από το διαθέσιμο σετ δεδομένων, χρησιμοποιείται, συνήθως ο αλγόριθμος του υπολογισμού της μέγιστης πιθανότητας (maximum likelihood estimation), όπου επιδιώκουμε να επιλέξουμε αυτές τις τιμές b που θα οδηγήσουν το μοντέλο σε προβλέψεις πολύ κοντά στο 1 για την κατηγορία που επιθυμούμε και, αντίστοιχα, κοντά στο 0 για την άλλη. Όπως και πριν, πρέπει και εδώ να χρησιμοποιηθεί ένας αλγόριθμος που θα πετύχει την ζητούμενη βελτιστοποίηση, έχοντας ως σημείο αναφοράς το σετ δεδομένων.

Εδώ, πρέπει ξανά να αναφέρουμε την ανάγκη για προ επεξεργασία του συνόλου δεδομένων, όπως την διαγραφή outlier τιμών και των εισόδων που παρουσιάζουν σημαντική γραμμική σχέση μεταξύ τους καθώς θα οδηγηθούμε σε overfitting .

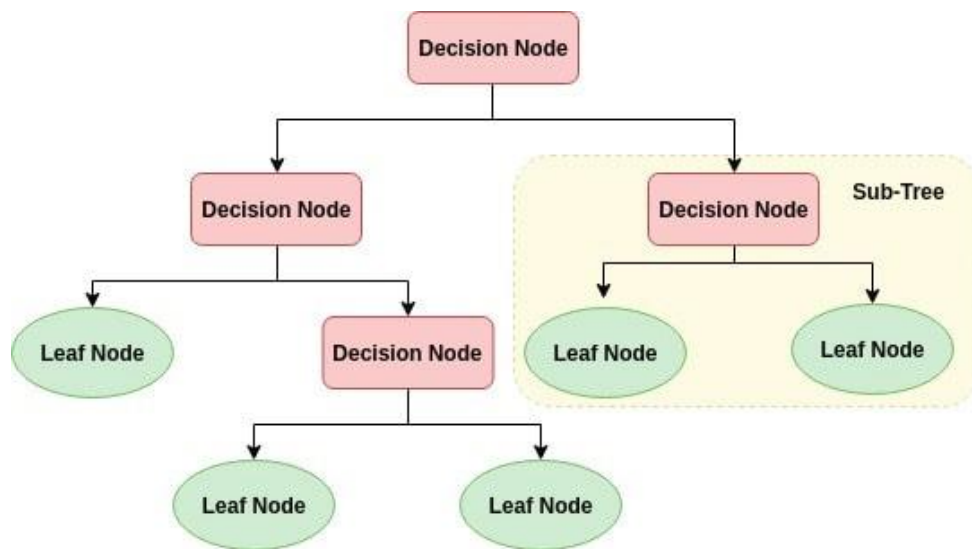
Αποφάσεις Μέσω Δένδρου (Decision Tree)

Η ονομασία του συγκεκριμένου τρόπου ανάλυσης οφείλεται στην αναπαράσταση δένδρου (Dev & Eden, 2019), που περιλαμβάνει το σύνολο των κανόνων-εντολών, σύμφωνα με τις οποίες γίνεται προσπάθεια να λυθεί το ζητούμενο πρόβλημα κατηγοριοποίησης/παλινδρόμησης. Ο στόχος, όπως αναφέρθηκε, είναι να κατασκευάσουμε ένα μοντέλο το οποίο αφού συμβουλευτεί συγκεκριμένους κανόνες που πηγάζουν από το σετ δεδομένων όπου έχει γίνει η εκπαίδευση, να παρέχει πρόβλεψη για την εξαρτημένη μεταβλητή.

Η μορφή του, όπως φαίνεται, και στην εικόνα 16, ομοιάζει αυτή του ανεστραμμένου δένδρου, όπου ένας εσωτερικός κόμβος (node) αντιπροσωπεύει κάποιο συγκεκριμένο

χαρακτηριστικό των δεδομένων που έχουμε, ενώ το κάθε ‘φύλλο’ μία από τις κατηγορίες που καλούμαστε να προβλέψουμε. Τέλος, έχουμε την ρίζα, δηλαδή τον αρχικό κόμβο.

Λόγω της δομής του συγκεκριμένου αλγόριθμου, είναι αρκετά εύκολο, υπό το πρίσμα του ανθρώπινου νου να γίνει κατανοητή η πορεία της σκέψης που οδηγεί στην εκτίμηση μίας κατηγορίας, αφού ουσιαστικά, αυτό που μας οδηγεί είναι η απάντηση με ένα ναι ή όχι σε ένα σύνολο ερωτήσεων, όπως φαίνεται παρακάτω.



Εικόνα 14. Αναπαράσταση δένδρου αποφάσεων

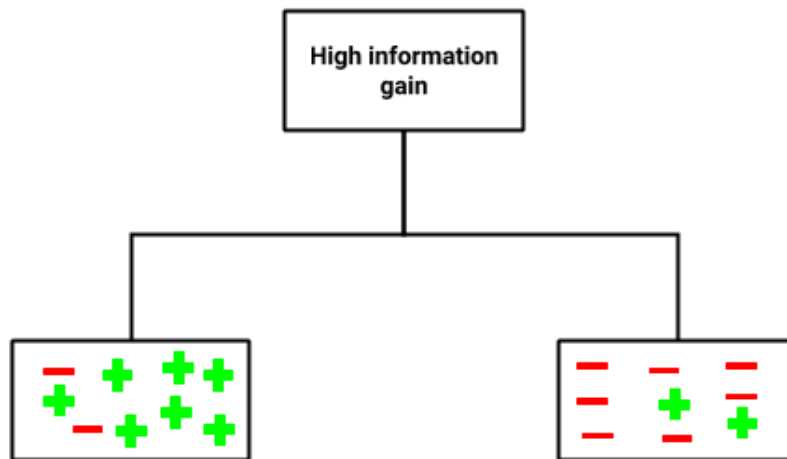
Ο τρόπος με τον οποίο καταλήγουμε στην κατασκευή ενός δένδρου παρουσιάζεται παρακάτω:

- Εύρεση και επιλογή του βέλτιστου feature ως αρχικό κόμβο μέσω της χρήσης κάποιας στατιστικής ανάλυσης/κριτηρίου
- Διαχωρισμός του ευρύτερου σετ δεδομένων σε μικρότερα υποσύνολα
- Επανάληψη της διαδικασίας μέχρι να συμβεί έστω μία από τις τρεις συνθήκες:
 - Δεν υπάρχουν άλλα δεδομένα
 - Δεν υπάρχουν άλλα διαθέσιμα attributes

Γίνεται εύκολα κατανοητό, βέβαια, ότι πρέπει να αναλυθούν τα συγκεκριμένα κριτήρια που λαμβάνονται υπόψιν όταν πραγματοποιείται η επιλογή των συγκεκριμένων nodes και να πραγματοποιηθεί μία σύντομη αναφορά στις βασικές τους έννοιες. Επιγραμματικά, αυτά είναι:

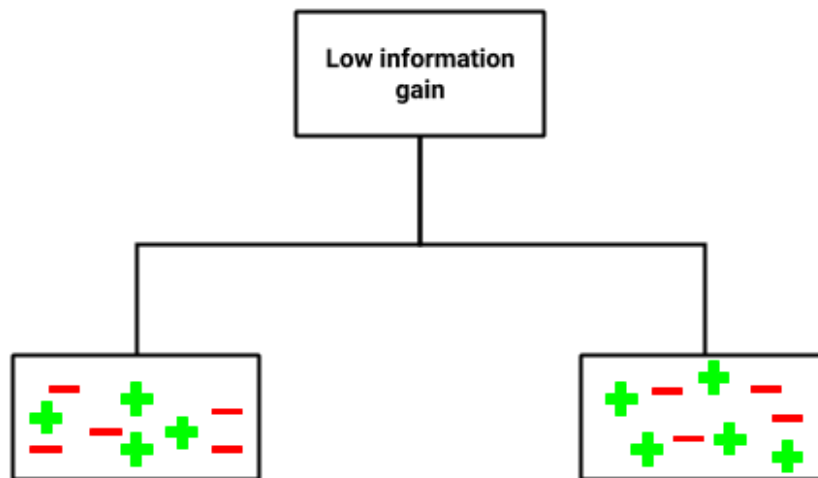
- **Κέρδος Πληροφορίας - Information Gain**
- **Gini Index**

Κέρδος Πληροφορίας (Information Gain) : Το κέρδος πληροφορίας αποτελεί μία στατιστική ιδιότητα που ουσιαστικά παρέχει γνώση σχετικά με το πόσο αποτελεσματικά το συγκεκριμένο χαρακτηριστικό είναι σε θέση να προχωρήσει σε διάκριση των διαφορετικών δεδομένων, υπό το πρίσμα της εκτίμησης μίας συγκεκριμένης κατηγορίας/τιμής. Παράδειγμα δύο αντιδιαμετρικά αντίθετων συστημάτων είναι τα παρακάτω:



Εικόνα 15. Σύστημα με υψηλό κέρδος πληροφορίας

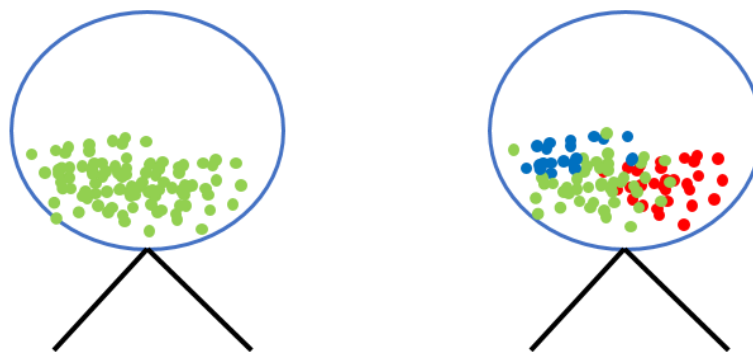
Μπορούμε στην εικόνα 15 πολύ εύκολα να δούμε ποια κατηγορία είναι η επικρατούσα σε κάθε κόμβο σε αντίθεση με την εικόνα 16.



Εικόνα 16. Σύστημα με χαμηλό κέρδος πληροφορίας

Για να είμαστε σε θέση να υπολογίσουμε, όμως την παραπάνω τιμή, πρέπει να αναφερθούμε και στην έννοια της Εντροπίας Πληροφορίας.

Εντροπία Πληροφορίας: Η Εντροπία Πληροφορίας (Entropy), εφευρέθηκε από τον Shannon και συμβολίζεται με το γράμμα H , στον τομέα της στατιστικής, αποτελεί το μέτρο της αβεβαιότητας μίας τυχαίας μεταβλητής ή αλλιώς, την περιγραφή της τυχαιότητας ενός συστήματος. Από τον ορισμό, είναι ευνόητο ότι αν θέλουμε να έχουμε δείγματα από μία μόνο κατηγορία ή τουλάχιστον σε σημαντικό ποσοστό, χρειαζόμαστε **μικρές** τιμές εντροπίας πληροφορίας. Σε περίπτωση που φυσικά αναφερόμαστε σε πλήρως ίδιο σύνολο, η τιμή της εντροπίας είναι μηδέν.



Εικόνα 17. Καθαρότητα ενός, υπό μελέτη, συστήματος

Gini Index: Σύμφωνα με το συγκεκριμένο μέγεθος, η πιθανότητα να επιλέξουμε, τυχαία δύο αντικείμενα ίδιας τάξης από έναν ‘καθαρό’ πληθυσμό, είναι 1. Σε περίπτωση όπου λάβουμε αυτό το κριτήριο υπόψιν στον σχηματισμό των δένδρων αποφάσεων, θα πρέπει αρχικά να υπολογίσουμε τους αντίστοιχους δείκτες για τους υπο-κόμβους αθροίζοντας το τετράγωνο της πιθανότητας επιτυχίας/αποτυχίας και στη συνέχεια, συγκρίνοντας τις τιμές ώστε να συμπεράνουμε ποιο χαρακτηριστικό είναι εκείνο με τον μεγαλύτερο δείκτη Gini.

Τέλος, θα πρέπει να σημειωθεί ότι η βελτιστοποίησή του διευκολύνεται, σημαντικά, κυρίως λόγω του μικρού αριθμού υπερ-παραμέτρων που πρέπει να οριστούν.

Παρόλο το γεγονός ότι η βελτιστοποίηση της συγκεκριμένης μεθόδου είναι εύκολη λόγω του μικρού αριθμού υπερ-παραμέτρων, η συγκεκριμένη μέθοδος παρουσιάζει και αρκετά μειονεκτήματα. Το κυριότερο εντοπίζεται στο γεγονός ότι είναι επιρρεπής σε ένα αρκετά διάσημο πρόβλημα της μηχανικής μάθησης, που θα αναλυθεί παρακάτω και ονομάζεται Overfitting. Σε αυτό έρχεται να προστεθεί και η παραγωγή μη αντικειμενικών εκτιμήσεων στις περιπτώσεις μη ισορροπημένου συνόλου δεδομένων (imbalanced) που θα αναλυθεί παρακάτω.

Naive Bayes

Η συγκεκριμένη τεχνική αποτελεί μία, αρκετά απλή αλλά διάσημη, μέθοδο για δυαδική κατηγοριοποίηση (Naive Bayes Classifier, 2018) βάση της οποίας είναι η υπόθεση της ανεξαρτησίας μεταξύ των διάφορων χαρακτηριστικών μέσα στο σύνολο δεδομένων (για αυτό και η ονομασία Naive – Αφελής.). Αυτό, σημαίνει ότι, σύμφωνα με την υπόθεσή μας, η αλλαγή της τιμής μίας μεταβλητής δεν επηρεάζει τις υπόλοιπες.

Για παράδειγμα, έστω ότι ένα σχολείο διαθέτει έναν πληθυσμό 50 ατόμων όπου αυτά τα άτομα, μπορούν να είναι είτε καθηγητές είτε μαθητές αλλά και μπορούν να είναι, είτε αρσενικό ή θηλυκό. Εμείς θέλουμε να υπολογίσουμε την πιθανότητα μίας γυναίκας να είναι μαθήτρια.

Με βάση τον αριθμό και τις αντίστοιχες ιδιότητες των ατόμων, είμαστε σε θέση να υπολογίσουμε την πιθανότητα της κάθε κατηγορίας, πχ αν έχουμε 20 γυναίκες τότε η

πιθανότητα να είναι γυναίκα είναι $2/5$, αν έχουμε 47 μαθητές, να είναι μαθητής $47/50$ κτλ.

Στην περίπτωση όπου θέλουμε να υπολογίσουμε την πιθανότητα ενός ατόμου, όπου εφόσον είναι γυναίκα να είναι μαθήτρια, θα πρέπει να συμβουλευτούμε τον γνωστό τύπο του Bayes για την υπο-συνθήκη ή δεσμευμένη πιθανότητα:

$$\Pr [A | B] = \frac{\Pr [A \cap B]}{\Pr [B]}.$$

Εικόνα 18. Υπό συνθήκη πιθανότητα

Σε σχέση με την κλασσική στατιστική όμως, και ειδικά στην περίπτωση που έχουμε περισσότερες μεταβλητές στην είσοδο, η υπόθεση της ανεξαρτησίας μας επιτρέπει να προχωρήσουμε στον πολλαπλασιασμό των πιθανοτήτων για κάθε μεταβλητή με την αντίστοιχη πιθανότητα να επιλέξουμε τυχαία την εκάστοτε οντότητα και το αποτέλεσμα αυτό να το διαιρέσουμε με το γινόμενο της συνολικής πιθανότητας να εμφανιστεί το εκάστοτε χαρακτηριστικό.

Για παράδειγμα, έστω ότι θέλουμε να διακρίνουμε με βάση τρία χαρακτηριστικά αν έχουμε επιλέξει πορτοκάλι, μπανάνα ή κάποιο άλλο φρούτο. Έστω ότι αυτά τα χαρακτηριστικά είναι το αν είναι πορτοκαλί, αν είναι στρογγυλό και αν είναι πικρό.

Υπολογίζοντας την πιθανότητα του κάθε φρούτου μέσα στον δεδομένο πληθυσμό-σετ δεδομένων (ονομάζεται και *prior* πιθανότητα), προχωράμε στον υπολογισμό, πρώτα του γινομένου των πιθανοτήτων ύπαρξης του κάθε χαρακτηριστικού(τιμή K), πχ πόσα φρούτα είναι στρογγυλά στον πληθυσμό και έπειτα το γινόμενο των δεσμευμένων πιθανοτήτων του κάθε χαρακτηριστικού για την κατηγορία που επιδιώκουμε να προβλέψουμε (*Probability of Likelihood* για το κάθε φρούτο).

Έτσι, μπορούμε, να προχωρήσουμε στην πρόβλεψη με βάση τα χαρακτηριστικά, για το πορτοκάλι, αντικαθιστώντας στην εξίσωση του Bayes τον αριθμητή και τον παρονομαστή με τις παραπάνω τιμές:

$$\frac{\text{prior (πορτοκάλι)} * \text{Probability of Likelihood(πορτοκάλι)}}{K}$$

Στην συγκεκριμένη μέθοδο μπορούμε να προχωρήσουμε σε βελτιώσεις όπως το να υπολογίσουμε την λογαριθμική πιθανότητα αλλά και να υποθέσουμε διαφορετικές διασπορές στα σύνολα δεδομένων μας.

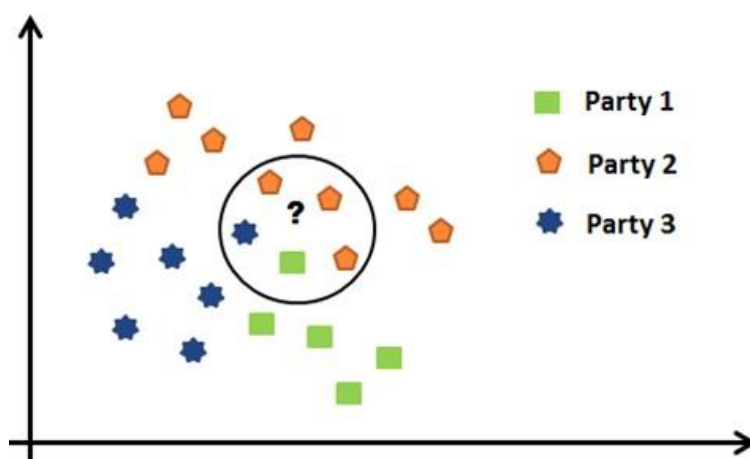
k - Nearest Neighbors (kNN)

Ο αλγόριθμος k αριθμού πλησιέστερων γειτόνων αποτελεί άλλον έναν αρκετά σημαντικό αλγόριθμο τόσο για παλινδρόμηση όσο και κατηγοριοποίηση. Η χρήση, όμως, εκ μέρους του, της ομοιότητας μεταξύ χαρακτηριστικών για την πρόβλεψη καινούργιων τιμών τον έχει μετατρέψει σε ένα σημαντικό εργαλείο για προβλήματα κατηγοριοποίησης στην βιομηχανία (Davies, 2005).

Η λογική με την οποία λειτουργεί, βασίζεται στον υπολογισμό της απόστασης (συνήθως ευκλείδεια) μεταξύ των δεδομένων διαφορετικών κατηγοριών και της, υπο εκτίμηση, μέτρησης. Με βάση αυτό το κριτήριο, στη συνέχεια πραγματοποιείται αύξουσα ταξινόμηση των προηγούμενων αποτελεσμάτων και αποφασίζεται με βάση τον αριθμό πλησιέστερων γειτόνων η κατηγορία της νέας τιμής.

Για παράδειγμα, έστω ότι θέλουμε να χρησιμοποιήσουμε αξιοποιήσουμε τους πλησιέστερους $k=5$ γείτονες για ένα συγκεκριμένο πρόβλημα κατηγοριοποίησης. Στην περίπτωση μας, δηλαδή αναφερόμαστε σε μία ανάλυση 5- Κοντινότερων Γειτόνων (kNN).

Έστω ακόμη ότι έχουμε 3 διαφορετικές κατηγορίες (Parties) με πράσινο, πορτοκαλί και μπλε χρώμα όπως φαίνεται και στην εικόνα 19.



Εικόνα 19 Πρόβλημα 5 κοντινότερων γειτόνων

Έστω, όπως βλέπουμε και στην εικόνα, ότι εμφανίζεται μία νέα, άγνωστη τιμή την οποία εμείς καλούμαστε να προβλέψουμε, την κατηγορία της.

Όπως μάθαμε, αρχικά, υπολογίζουμε τις ευκλείδειες αποστάσεις από τα σημεία του υπάρχοντος σετ δεδομένων και τις ταξινομούμε σε αύξουσα σειρά, διατηρώντας τις πρώτες 5 (αυτή είναι η τιμή k που έχουμε επιλέξει εδώ). Έπειτα, με βάση την κατηγορία που πλειοψηφεί στο σύνολο των 5 αυτών γειτόνων επιλέγουμε αντίστοιχα να ταξινομήσουμε την καινούργια άγνωστη τιμή.

Ο συγκεκριμένος αλγόριθμος, όπως φαίνεται συνδυάζει την απλότητα με ισχυρή ικανότητα προβλέψεως, με συγκεκριμένους, όμως περιορισμούς. Αυτοί εντοπίζονται στις περιπτώσεις όπου το σύνολο των δεδομένων ή/και των χαρακτηριστικών είναι σημαντικά μεγάλα όπου προκαλεί την αντίστοιχη χρήση μεγάλου αριθμού υπολογιστικών πόρων.

Support Vector Machines

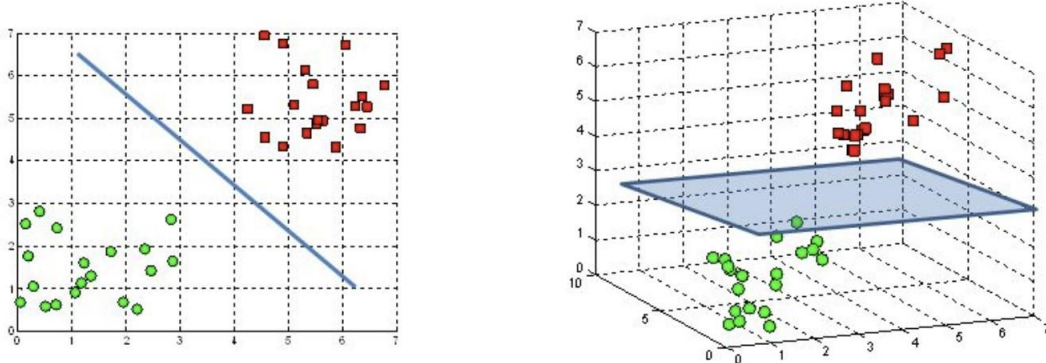
Στο ίδιο μοτίβο με τον αλγόριθμο k κοντινότερων γειτόνων οι Μηχανές Διανυσμάτων Υποστήριξης ή αλλιώς, επίσημα, Support Vector Machines παρουσιάζουν, προτίμηση, από τους ερευνητές για θέματα, κυρίως κατηγοριοποίησης στην σημερινή βιομηχανία (Pisner & Schnyer, 2020).

Διερευνώντας τον βασικό τρόπο λειτουργίας, μπορούμε να συμπεράνουμε ότι μία SVM ανάλυση επιδιώκει στην εύρεση ενός βέλτιστου υπερ-επιπέδου ή αλλιώς hyperplane το οποίο θα είναι σε θέση να διακριτοποιήσει, με μέγιστο δυνατό περιθώριο τα δεδομένα του, υπό μελέτη, συνόλου.

Αρχικά, προκειμένου να είμαστε σε θέση να κατανοήσουμε επαρκώς τον αλγόριθμο πρέπει να αναλυθεί η έννοια του **υπερ-επιπέδου**. Το υπερ επίπεδο είναι ένα όριο απόφασης του οποίου η γεωμετρική μορφή εξαρτάται από τις διαστάσεις των χαρακτηριστικών του συνόλου δεδομένων. Το ίδιο είναι υπεύθυνο για την κατηγοριοποίηση των διαφορετικών σημείων και για αυτό το λόγο η εύρεση του βέλτιστου αποτελεί το σημαντικότερο ζήτημα σε μία SVM ανάλυση. Συγκεκριμένα,

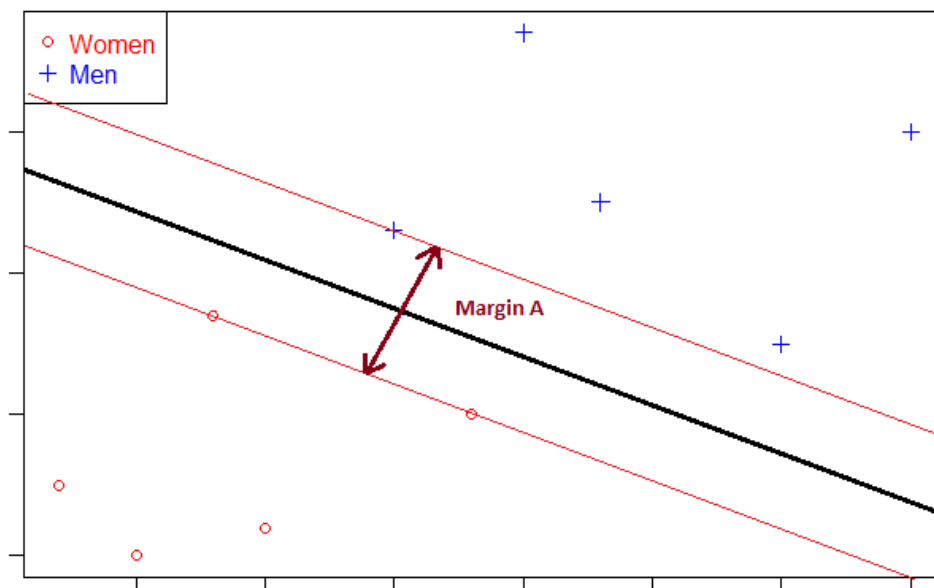
λαμβάνοντας υπόψιν την θέση και την μορφή του είμαστε σε θέση απευθείας να ορίσουμε την κατηγορία της καινούργιας-άγνωστης μεταβλητής

Για παράδειγμα, στην απλή περίπτωση που υπάρχουν μόνο 2 χαρακτηριστικά, η μορφή του είναι αυτή της γραμμής, ενώ σε 3, ενός επιπέδου, όπως φαίνεται παρακάτω



Εικόνα 20. Γεωμετρική μορφή του υπερ επιπέδου

Το βασικό πρόβλημα, όπως είναι λογικό εντοπίζεται στο πώς θα μπορέσουμε να ορίσουμε το συγκεκριμένο επίπεδο καθώς και τον τρόπο που θα βρούμε το βέλτιστο δυνατό. Για ένα τυχαίο υπερ επίπεδο, όπως στο παρακάτω σχήμα, το μέτρο της τιμής του περιθωρίου, που προαναφέραμε, ορίζεται ως η απόσταση μεταξύ του υπερ επιπέδου και του κοντινότερου σημείου του συνόλου δεδομένων.



Εικόνα 21. Τιμή του Margin για τυχαίο υπερ επίπεδο

Εκτελώντας διανυσματικές πράξεις μεταξύ των πλησιέστερων σημείων (που ονομάζονται και Διανύσματα Υποστήριξης) και έχοντας ως οδηγό τις τιμές των

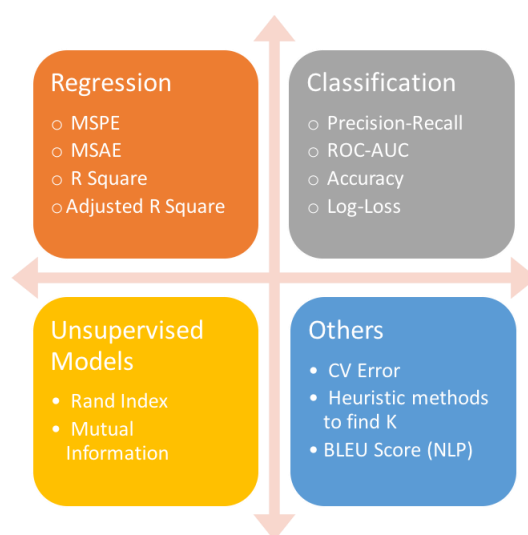
αποστάσεων μεταξύ αυτών και του υπερεπιπέδου, είμαστε σε θέση να βρούμε την βέλτιστη δυνατή μορφή που παρέχει το μέγιστο δυνατό περιθώριο.

Η χρήση της συγκεκριμένης τεχνικής, είναι δυνατόν να μας παρέχει σημαντική εκτιμητική ικανότητα στις περιπτώσεις όπου διαθέτουμε προβλήματα κατηγοριοποίησης με πολλά χαρακτηριστικά ενώ, ταυτόχρονα παρουσιάζεται αποδοτική αξιοποίηση των υπολογιστών πόρων .

Όπως, όμως έχουμε δει και σε προηγούμενους αλγορίθμους, έτσι και εδώ είναι σημαντική η αποφυγή χρήσης imbalanced σετ δεδομένων ενώ επιπλέον η αποδοτικότητά του ενδέχεται να υποφέρει σε προβλήματα μεγάλων δεδομένων (Big Data Analysis).

Δείκτες Αξιολόγησης

Οι δείκτες αξιολόγησης της εκάστοτε ανάλυσης μηχανικής μάθησης και η παρακολούθησή τους από τους ερευνητές αποτελεί την πιο σημαντική διαδικασία, τόσο για την αξιολόγηση των αποτελεσμάτων-προβλέψεων που έχουν πραγματοποιηθεί όσο και τις δυνατότητες και πρακτικές βελτιστοποίησης ορισμένων υπερ παραμέτρων του εκάστοτε μοντέλου. Στην συγκεκριμένη ενότητα θα αναλύσουμε ορισμένους από τους πιο βασικούς, που όπως θα δούμε αρμόζουν έκαστοι σε συγκεκριμένες κατηγορίες προβλημάτων, κυρίως λόγω της φύσης του καθενός.



Εικόνα 22. Δείκτες – Metrics Ανάλογα το πρόβλημα που μελετάται

Confusion Matrix

Στις περιπτώσεις όπου αντιμετωπίζουμε πρόβλημα επιβλεπόμενης μάθησης ένα ισχυρό εργαλείο αξιολόγησης της ικανότητας πρόβλεψης ενός μοντέλου είναι ο Confusion Matrix (Kulkarni & Batarseh, 2020).

Η βάση της λειτουργίας του, στηρίζεται στο ο ίδιος αποτελεί έναν πίνακα με 4 διαφορετικούς συνδυασμούς των εκτιμημένων και πραγματικών τιμών.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

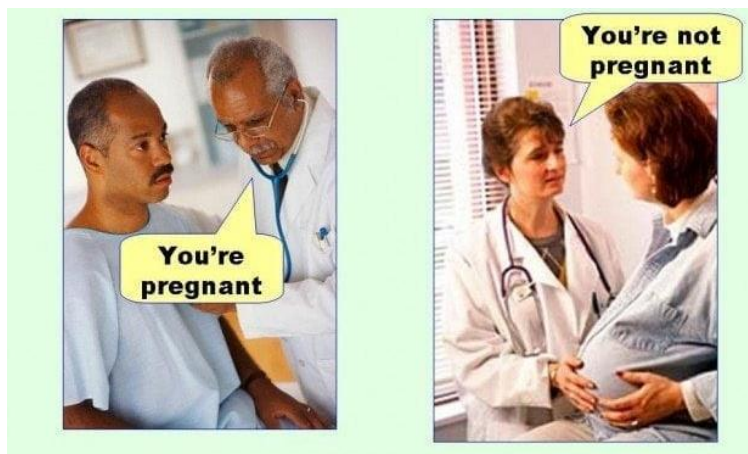
Εικόνα 23. Παράδειγμα Confusion Matrix

Για να κατανοήσουμε την σημασία του αλλά και να προχωρήσουμε στο πώς μέσω αυτού μπορούμε να υπολογίσουμε και άλλες σημαντικές στατιστικές ιδιότητες – metrics θα πρέπει να δούμε τι σημαίνουν τα ακρωνύμια TP, FP, FN και FN.

Αναλυτικά :

- **Αληθές Θετικό - True Positive (TP)** : Αφορά την αναγνώριση μίας κατάστασης-κλάσης, στην περίπτωση όπου πράγματι εκείνη συμβαίνει.
- **Αληθές Αρνητικό - True Negative (TN)** : Αφορά την αναγνώριση απουσίας μίας κατάστασης, όταν εκείνη, πράγματι, δεν συμβαίνει.
- **Ψευδές Θετικό - False Positive (FP)** : Αφορά την *λανθασμένη* αναγνώριση μίας κατάστασης, όταν εκείνη δεν συμβαίνει στην πραγματικότητα.
- **Ψευδές Αρνητικό - False Negative (FN)** : Αφορά την *λανθασμένη* αναγνώριση απουσίας μίας κατάστασης, όταν εκείνη, συμβαίνει στην πραγματικότητα.

Για την καλύτερη κατανόηση των παραπάνω όρων, παρουσιάζεται ένα γλαφυρό παράδειγμα ψευδούς θετικής και ψευδούς αρνητικής πρόβλεψης, στην εικόνα 24.



Εικόνα 24. *False Positive και False Negative Περιπτώσεις*

Πάνω στις συγκεκριμένες μεταβλητές μπορούμε να ορίσουμε συγκεκριμένα μεγέθη που μας δίνουν μία ευρύτερη εικόνα για την ικανότητα πρόβλεψης του εκτιμητή την *Specificity* και την *Sensitivity*.

Sensitivity

Αποτελεί το μέτρο για την ικανότητα του κατηγοροποιητή να εντοπίσει – αντιληφθεί επιτυχώς τα αληθώς θετικά αποτελέσματα και, όπως είναι λογικό σχετίζεται με τον αριθμό των, αναγνωρισμένων, θετικά αντικειμένων ως προς το σύνολο των θετικών δεδομένων.

Specificity

Αντίστοιχα, εδώ παρουσιάζεται η ικανότητα του κατηγοροποιητή να εντοπίζει τις αρνητικές τιμές και, βέβαια, εδώ το μέτρο είναι ανάλογο του κλάσματος του αριθμού των αρνητικά αναγνωρισμένων αντικειμένων ως προς του σύνολο των δειγμάτων.

Στην περίπτωση της πραγματοποίησης στατιστικών υποθέσεων, υπάρχουν δύο μεγέθη που σχετίζονται με την απόρριψη ή αποδοχή τους, τα οποία ονομάζονται Τύποι Λάθους 1 και 2. Συγκεκριμένα:

Τύποι I Λάθους

Ο συγκεκριμένος δείκτης μετρήσεων, που βρίσκεται και στην βιβλιογραφία ως FPR (False Positive Rate) αφορά το πλήθος των αναγνωρισμένων, λανθασμένα θετικών False Positive δειγμάτων, προς το σύνολο των αρνητικών.

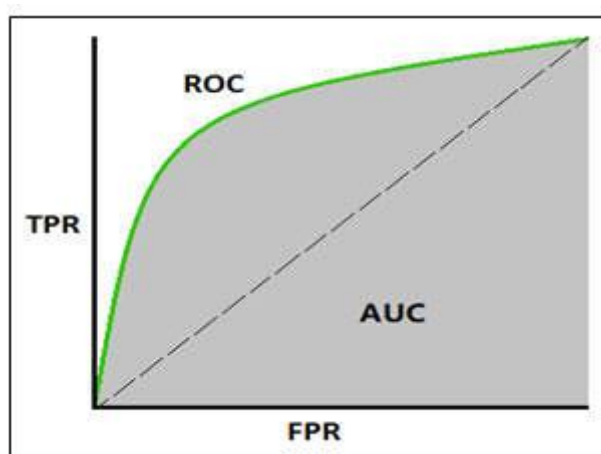
Τύπος II Λάθους

Εδώ, σε αντίθεση με πριν μας ενδιαφέρει ο αριθμός των δειγμάτων που έχουν αναγνωριστεί ως ψευδώς αρνητικά ως προς το σύνολο των θετικών.

Συνδυάζοντας τις παραπάνω έννοιες μπορούμε να προχωρήσουμε και σε γραφικές αναπαραστάσεις της ικανότητας πρόβλεψης του εκάστοτε κατηγοροποιητή. Οι συγκεκριμένοι δείκτες ονομάζονται ROC και AUC.

ROC/AUC

Στην περίπτωση τυχαίας επιλογής μίας θετικής και μίας αρνητικής κατάστασης, το μέτρο τους είναι σε θέση να μας παρέχει την πιθανότητα η πρώτη, να είναι μεγαλύτερη από την δεύτερη. Όπως, μπορεί να φανεί και στο σχήμα ο υπολογισμός της περιοχής κάτω από την καμπύλη, στο επίπεδο, με τον δείκτη FPR στον άξονα τον x ή αλλιώς και Τύπο I λάθους που αναφέραμε και αντίστοιχα στον y το True Positive Rate (TRP) μας παρέχει την AUC τιμή.



Εικόνα 25. ROC – AUC καμπύλες

Η τιμή τους κυμαίνεται, αφού μιλάμε για πιθανότητα, μεταξύ του 0 και 1 και, σε περίπτωση όπου δεν έχουμε imbalanced σετ δεδομένων θα πρέπει η τιμή να προσεγγίζει το 0.5.

Επίσης, αρκετά σημαντικές έννοιες που πρέπει να αναλυθούν, έστω συνοπτικά αφορούν τις έννοιες του *Accuracy* και του *Precision*.

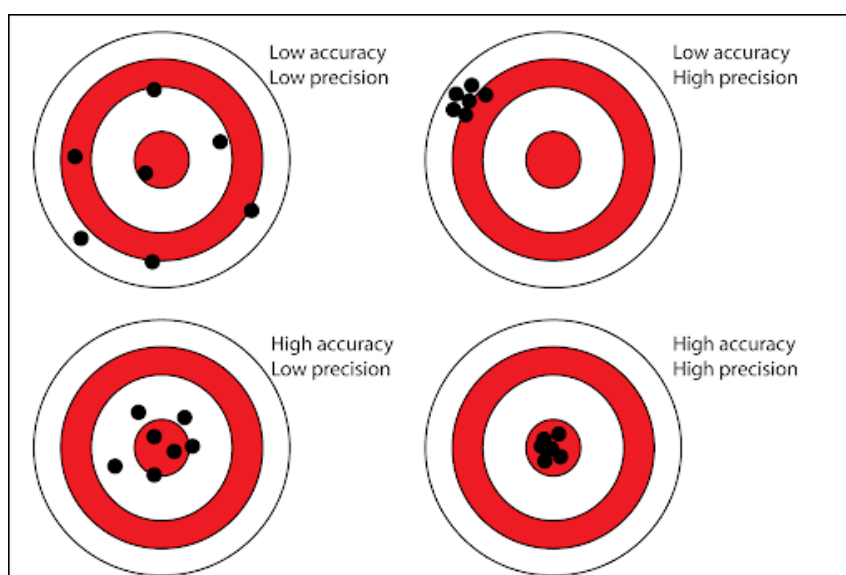
Ακρίβεια - Accuracy

Η ακρίβεια - Accuracy, είναι υπεύθυνη για τον υπολογισμό των δειγμάτων που έχουν ταξινομηθεί-κατηγοριοποιηθεί σωστά.

Precision

Το precision, από την άλλη σαν όρος, αφορά το πόσο κοντά βρίσκονται οι εκτιμήσεις μεταξύ τους.

Στην πλειονότητα των περιπτώσεων επιθυμούμε τα συστήματά μας να διαθέτουν υψηλές τιμές Precision και Accuracy



Εικόνα 26. Accuracy - Prediction

Λοιπές Έννοιες Μηχανικής Μάθησης

Πριν προχωρήσουμε είναι αρκετά σημαντικό να γίνει μία σύντομη καταγραφή ορισμένων βασικών εννοιών στον τομέα της μηχανικής μάθησης που έχουν ήδη αναφερθεί έως τώρα επιγραμματικά, των οποίων όμως η κατανόηση συνεισφέρει σημαντικά στην εκτέλεση εκ βάθους και αποτελεσματικών αναλύσεων.

Μη ισορροπημένο σετ δεδομένων - Imbalanced Dataset

Η ποιότητα του σετ δεδομένων τόσο για το στάδιο της εκπαίδευσης όσο και του validation, αποτελεί πλήρως καθοριστικό παράγοντα για την παραγωγή αξιόπιστων και χρήσιμων κατηγοροποιητών. Σε περίπτωση όπου έχουμε ένα πρόβλημα αναγνώρισης πολλαπλών κατηγοριών, είναι σημαντικό κάθε κλάση να αναπαρίσταται επαρκώς μέσα στο σετ δεδομένων. Υπάρχουν όμως και παραδείγματα όπου το πρόβλημα του imbalanced σετ δεδομένων είναι σχεδόν, αδύνατον να αποφευχθεί. Τέτοια παραδείγματα μπορούν να βρεθούν τόσο στον χρηματοπιστωτικό τομέα όσο και στην βιοϊατρική.

Στον χρηματοπιστωτικό τομέα, ένα βασικό παράδειγμα αποτελεί ο εντοπισμός απάτης σε συναλλαγές καθώς το ποσοστό τους σε σύγκριση με τις έγκυρες περιπτώσεις είναι σημαντικά μικρότερο. Στον τομέα της ιατρικής, η αναγνώριση μίας εξαιρετικά σπάνια αρρώστιας, αφού κατά τον ορισμό οι άνθρωποι που θα την έχουν είναι μία σημαντική μειοψηφία.

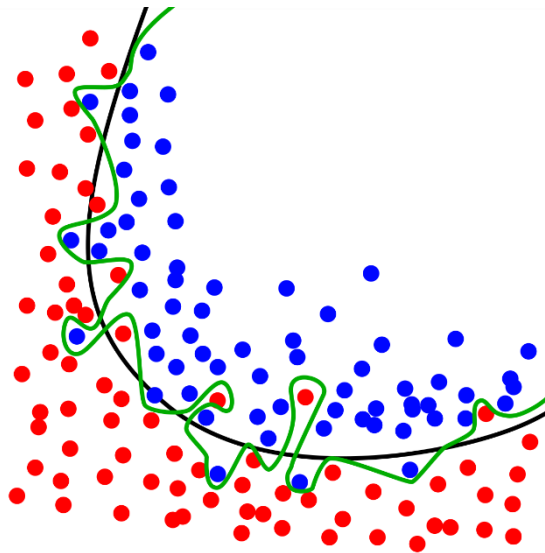
Στις περιπτώσεις αυτές, προτείνεται ένα σύνολο διαφορετικών τεχνικών με σκοπό την πανάκεια της φύσης του ζητήματος όπως αλλαγή του αλγορίθμου μάθησης που αξιοποιείται από τον κατηγοροποιητή, των δεικτών αξιολόγησης ενώ και προσπάθειες επαναδειγματοληψίας ρίχνοντας βάρος στην προσθήκη αντιγράφων της κατηγορίας με την μικρότερη παρουσία στα δεδομένα, ειδικά αν δεν είμαστε σε θέση να βρούμε, εκ φύσεως περισσότερα.

Υπερπροσαρμογή - Overfitting

Η υπερπροσαρμογή (Overfitting) αποτελεί ένα από τα πιο συχνά προβληματικά φαινόμενα στον τομέα της μηχανικής μάθησης (Simon, 2013).

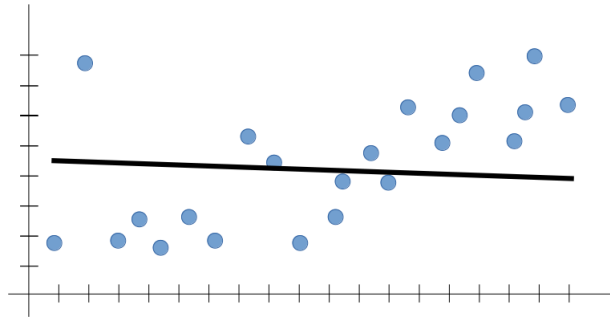
Συγκεκριμένα, όταν συμβαίνει αυτό το φαινόμενο, το μοντέλο κατηγοριοποίησης μπορεί να παρουσιάσει υψηλούς δείκτες σχετικά με την ικανότητα επιτυχημένης πρόβλεψης δεδομένων σε ένα γνωστό σετ, αλλά και την φανερή αδυναμία να πραγματοποιήσει προβλέψεις στις περιπτώσεις όπου έρθουν καινούργια – άγνωστα δεδομένα.

Για παράδειγμα, στην παρακάτω εικόνα, παρατηρούνται δύο προσεγγίσεις – γραμμές . Ενώ η μία, στο γνωστό σετ δεδομένων θα παρουσιάσει αυξημένους δείκτες επιτυχίας, δεν θα μπορέσει, κατά πάσα πιθανότητα να γενικεύσει αυτά τα αποτελέσματα στις περιπτώσεις καινούργιων δεδομένων, σε σύγκριση με την δεύτερη καμπύλη.



Εικόνα 27. Overfitting

Τέλος, ακριβώς το αντίθετο φαινόμενο εντοπίζεται στη αποτυχία κατασκευής μίας προσεγγιστικής συνάρτησης ταξινόμησης για το δεδομένο σετ. Επισήμως, ονομάζεται Underfitting και συχνά, είναι ήδη έκδηλο από τους χαμηλούς δείκτες αξιολόγησης του κατηγοροποιητή



Εικόνα 28. Φαινόμενο Underfitting

Μηχανική μάθηση vs βαθιά μάθηση

Τι είναι βαθιά μάθηση

Η βαθιά μάθηση (deep learning) είναι ένας επιστημονικός κλάδος μια μεγαλύτερης οικογένειας της μηχανικής όρασης. Οι μέθοδοι της βασίζονται στην εκμάθηση των δεδομένων των αναπαραστάσεων, σε αντίθεση με τη μηχανική μάθηση που ζητείται συγκεκριμένη χρήση αλγόριθμων. Η βαθιά μάθηση (βαθιά δομημένη μάθηση, ιεραρχική μάθηση ή βαθιά μηχανική μάθηση) επιχειρεί σε επίπεδα επεξεργασίας με σύνθετες δομές ή αλλιώς αποτελείται από πολλαπλούς μη γραμμικούς μετασχηματισμούς. Η βαθιά μάθηση είναι μέρος μιας ευρύτερης οικογένειας των μεθόδων μηχανικής μάθησης που βασίζονται σε μαθησιακές αναπαραστάσεις δεδομένων. Μια παρατήρηση ή ένα δεδομένο μπορεί να αναπαρασταθεί με πολλούς διαφορετικούς τρόπους. Ορισμένες παραστάσεις διευκολύνουν την εκπαίδευση ενώ άλλες λόγω των δυνατοτήτων της βαθιάς μάθησης να αντικαθιστά χειροποίητα χαρακτηριστικά με αποτελεσματικούς αλγόριθμους για μη εποπτευόμενη ή ημι-εποπτευόμενη εκμάθηση, βοηθούν στην αντίστοιχη εξαγωγή τους.

Υπάρχουν σημαντικά πολλές κατηγορίες – αρχιτεκτονικές διατάξεις των νευρωνικών δικτύων όπως :

- Συνεχή βαθιά νευρωνικά δίκτυα
- Βαθιά δίκτυα πεποιθήσεως
- Επαναλαμβανόμενα νευρωνικά δίκτυα

Τα οποία γνωρίζουν άνηση σε πληθώρα εφαρμογών όπως υπολογιστική όραση, αυτόματη αναγνώριση ομιλίας, επεξεργασία φυσικής γλώσσας αναγνώριση ήχου και

σημάτων, εφαρμογές βιοπληροφορική με αξιοσημείωτα αποτελέσματα στα παραπάνω.

Εναλλακτικά, η βαθιά μάθηση έχει χαρακτηριστεί ως λέξη-κλειδί, ή μετονομασία νευρωνικά δίκτυα Η βαθιά μάθηση θα μπορούσε να χαρακτηριστεί ως μια κατηγορία αλγορίθμων μηχανικής μάθησης που χρησιμοποιούν πολλά επίπεδα μη γραμμικών μονάδων επεξεργασίας για εξαγωγή χαρακτηριστικών. Κάθε διαδοχικό επίπεδο χρησιμοποιεί την έξοδο από το προηγούμενο επίπεδο ως είσοδο.

Οι αλγόριθμοι μπορεί να επιβλέπονται ή να μην ελέγχονται και οι εφαρμογές περιλαμβάνουν ανάλυση προτύπων (χωρίς επίβλεψη) και ταξινόμηση (υπό επίβλεψη). Βασίζονται στην (χωρίς επίβλεψη) εκμάθηση πολλαπλών επιπέδων χαρακτηριστικών ή αναπαραστάσεων του δεδομένα. Τα χαρακτηριστικά υψηλότερου επιπέδου προέρχονται από χαρακτηριστικά χαμηλότερου επιπέδου για να σχηματίσουν μια ιεραρχική αναπαράσταση. Είναι μέρος του ευρύτερου πεδίου μηχανικής μάθησης των μαθησιακών αναπαραστάσεων δεδομένων.

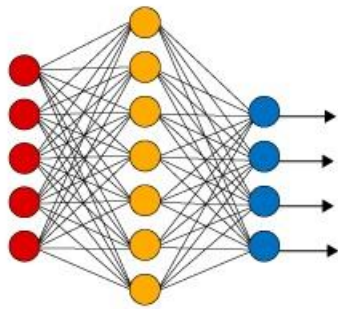
Τα νευρικά δίκτυα λαμβάνουν τις αναπαραστάσεις τους από τη χρήση επιπέδων μάθησης. Οι πρωτεύοντες εγκέφαλοι κάνουν ένα παρόμοιο πράγμα στον οπτικό φλοιό, οπότε η ελπίδα ήταν ότι η χρήση περισσότερων στρωμάτων σε ένα νευρικό δίκτυο θα μπορούσε αφήσει να μάθει καλύτερα μοντέλα. Ωστόσο, μελέτες έχουν δείξει ότι οι εσωτερικές αναπαραστάσεις μεταξύ αυτών των μοντέλων δεν μπόρεσαν να λειτουργήσουν, αλλά ωστόσο, τα επιτυχημένα μοντέλα συνειδητοποιήθηκαν ότι θα κατασκευάσουν ένα ρηχό δίκτυο, ένα με ένα μόνο επίπεδο αναπαράστασης δεδομένων Μαθαίνοντας σε βάθος νευρωνικό δίκτυο, ένα με περισσότερα από ένα επίπεδα αναπαράστασης δεδομένων, απλά δεν λειτούργησε.

Η ιδέα της βαθιάς μάθησης ξεκίνησε με την πτώση των τιμών των υλικών στον τομέα τον υπολογιστών αλλά και με την κατακόρυφη ανάπτυξη των γραφικών καρτών (GPU). Κύριος στόχος της είναι να αντικατοπτρίσει σ' ένα πρότυπο μοντέλου, τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος επεξεργάζεται το φως και τον ήχο και τον μετατρέπει σε όραση και ακοή. Ουσιαστικά είναι ένας τρόπος εκμάθησης των υπολογιστών ώστε να κατανοήσουν το πως οι άνθρωποι αντιδρούν φυσιολογικά, βάση την εμπειρία και τον παραδειγματισμό κατά τη διάρκεια της ζωής τους.

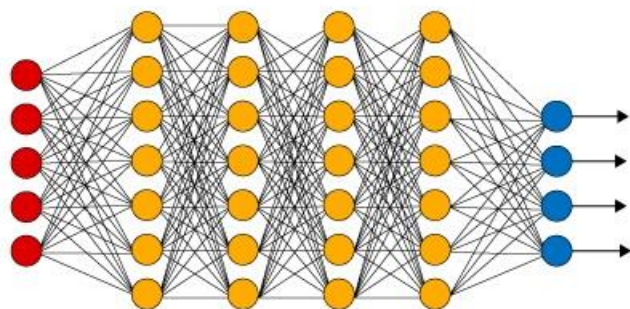
Η βαθιά μάθηση είναι μια τεχνική της μηχανικής μάθησης, στην οποία τα χαρακτηριστικά και οι διεργασίες μαθαίνονται απευθείας από τα δεδομένα που υπάρχουν. Τέτοια δεδομένα μπορεί να είναι αρχεία εικόνας, κειμένου ή ήχου. Τα μοντέλα της βαθιάς μάθησης μπορούν να αποδώσουν ένα απίστευτα μεγάλο ποσοστό εγκυρότητας, που πολλές φορές ξεπερνάει ακόμα και την απόδοση του ανθρώπινου εγκεφάλου. Αυτά τα μοντέλα εξοικειώνονται με τη χρήση μεγάλου όγκου ενδεικτικών δεδομένων, αλλά και αρχιτεκτονικών των νευρωνικών δικτύων που παρέχουν πολλά επίπεδα. Ένα πασίγνωστο μοντέλο νευρωνικού δικτύου στη βαθιά μάθηση είναι το CNN (Convolutional Neural Network), το οποίο δουλεύει πολύ καλά με τη χρήση αρχείων εικόνων.

Ένα πλήθος μελετών σε αυτό τον τομέα επιδιώκουν να πραγματοποιήσουν καλύτερες και βέλτιστες αναπαραστάσεις με σκοπό να δημιουργηθούν μοντέλα τα οποία θα μάθουν από τα συγκεκριμένα δεδομένα μεγάλης κλίμακας με ή και χωρίς σημάνσεις. Μερικές από τις συγκεκριμένες παραστάσεις προέρχονται από καινούργια τεχνολογικά δεδομένα και έρευνες σε επιστήμες όπως η νευροεπιστήμη και η ιατρική ενώ ταυτόχρονα βασίζουν τον κορμό της λειτουργίας τους στα πρότυπα και αρχιτεκτονικές επικοινωνίας και επεξεργασίας πληροφοριών σε νευρικά συστήματα όπως οι κωδικοποιήσεις που πραγματοποιώντας και οι αντίστοιχες σχέσεις ενεργοποίησης νευρικών κυττάρων και ερεθισμάτων του εγκεφάλου. Εκτός του ότι η αρχιτεκτονική της βαθιάς μάθησης μπορεί να συσχετιστεί και με την επεξεργασία πληροφοριών και επικοινωνιακών προτύπων σ' ένα βιολογικό νευρωνικό δίκτυο, όπως η νευρωνική κωδικοποίηση που επιχειρεί τον καθορισμό της σχέσης μεταξύ διαφόρων ερεθισμάτων και συνδεδεμένων νευρωνικών αποκρίσεων στον εγκέφαλο, η αρχιτεκτονική των βαθιά νευρωνικών δικτύων (deep neural networks), βαθιών δικτύων πίστης (deep belief networks) αλλά και των επαναλαμβανόμενων νευρωνικών δικτύων (recurrent neural networks) έχουν εφαρμοστεί στα πεδία που περιλαμβάνουν την μηχανική όραση, την αναγνώριση ομιλίας, την επεξεργασία της φυσικής γλώσσας, την αναγνώριση ήχου, το φιλτράρισμα των μέσων κοινωνικής δικτύωσης, τη μετάφραση στη γλώσσα των μηχανών, τη βιοπληροφορική αλλά και το σχεδιασμό φαρμάκων. Τα αποτελέσματα αυτών των εφαρμογών καθιστούν τις μηχανές ισάξιες έως και ανώτερες βάση αποτελεσμάτων ακόμα και των πιο ειδικών.

Simple Neural Network



Deep Learning Neural Network



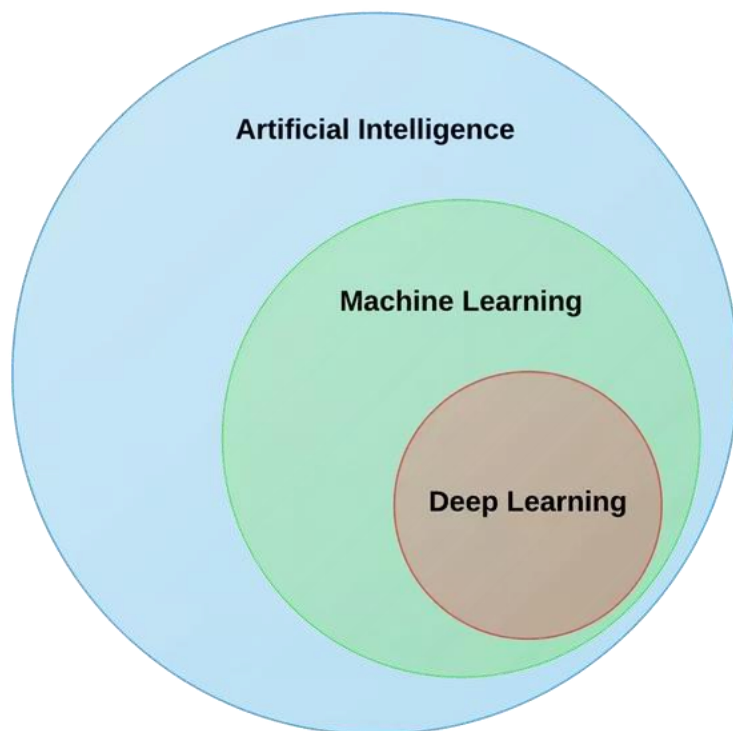
● Input Layer ● Hidden Layer ● Output Layer

Εικόνα 29. στα αριστερά παρατηρούμε ένα απλό νευρωνικό δίκτυο και δεξιά ένα νευρωνικό δίκτυο της βαθιάς μάθησης

Διάφορες μεταξύ μηχανικής και βαθιάς μάθησης

Η μηχανική και η βαθιά μάθηση προσφέρουν ποικίλους τρόπους στην εκπαίδευση μοντέλων και την ταξινόμηση δεδομένων. Έχουν όμως και σημαντικές διαφορές που τις κάνουν να ξεχωρίζουν η μία από την άλλη.

Η μεταφορά της αναγνώρισης και της εξοικείωσης του ανθρώπινου περιβάλλοντος σε μηχανικό σύστημα είναι ο πρωτεύον στόχος και των δύο αυτών πεδίων. Η εκμάθηση και η αναγνώριση παραδειγμάτων σε άριστο ποσοστό επιτυχίας είναι το ζητούμενο που ερευνάται στα δύο πεδία αλλά ο τρόπος που διεξάγεται στο καθένα είναι ξεχωριστός.

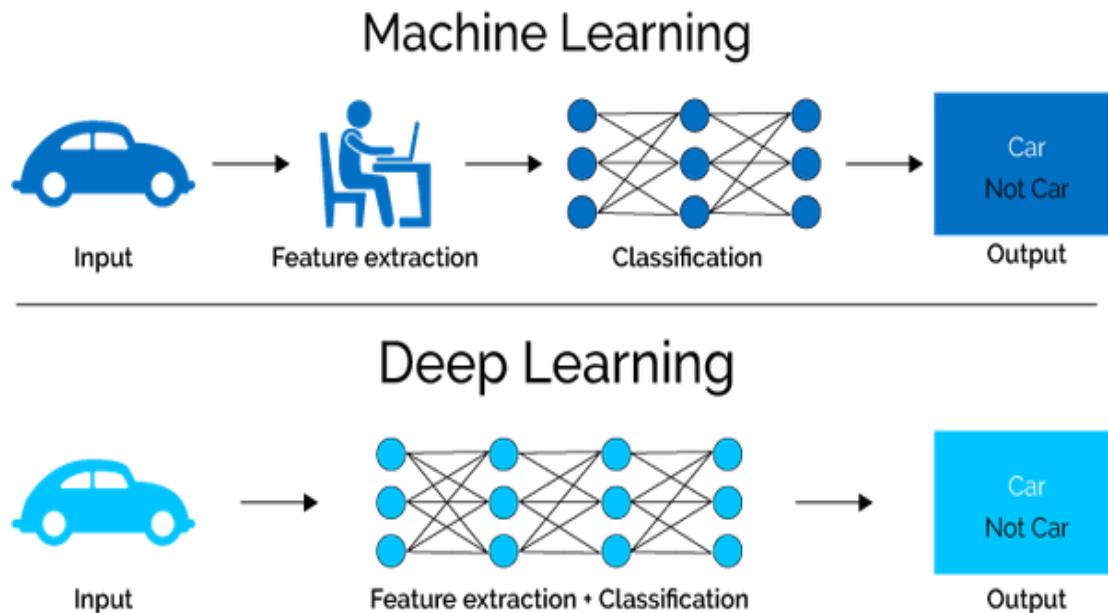


Εικόνα 30. Η συσχέτιση μεταξύ της τεχνητής νοημοσύνης, της μηχανικής μάθησης και της βαθιάς μάθησης

Χρησιμοποιώντας έναν υπολογιστή για ταξινόμηση εικόνων με την προσέγγιση της μηχανικής μάθησης, η επιλογή των βασικών χαρακτηριστικών, όπως των άκρων και των γωνιών του αντικειμένου που υπάρχει στην εικόνα, γίνεται με τη βοήθεια ανθρώπινου παράγοντα. Αυτό αποσκοπεί στην εκμάθηση του μοντέλου με βάση τη μηχανική μάθηση. Έπειτα, το συγκεκριμένο μοντέλο κάνει μια αναφορά αναλύοντας και ταξινομώντας τα νέα αντικείμενα. Αυτή η τεχνική γίνεται επίσης για την αναγνώριση σκηνης ή τοπίου και την ανίχνευση αντικειμένων. Η λύση προβλημάτων με τη μέθοδο της μηχανικής μάθησης έχει συγκεκριμένη ροή ενεργειών. Παρατηρώντας την εικόνα, γίνεται εξαγωγή των βασικών στοιχείων που τη διέπουν και στη συνέχεια δημιουργείται ένα μοντέλο το οποίο περιγράφει ή υποθέτει το αντικείμενο.

Αντιθέτως, στη βαθιά μάθηση παραλείπεται ο ανθρώπινος παράγοντας ώστε να γίνει η εξαγωγή των χαρακτηριστικών των εικόνων. Αντ' αυτού, γίνεται άμεση εισαγωγή της εικόνας στον αλγόριθμο της βαθιάς μάθησης που αποτελείται το σύστημα, και τότε γίνεται η πρόβλεψη του αντικειμένου. Έτσι, η βαθιά μάθηση κηρύσσεται ως

έναν υπό τύπος της μηχανικής μάθησης, διότι επεξεργάζεται απευθείας τις εικόνες και συνήθως είναι πιο περίπλοκη από τη μηχανική μάθηση, λόγω των χαρακτηριστικών που την απαρτίζουν.



Εικόνα 31. Η βασική διαφορά μεταξύ μηχανικής και βαθιάς μάθησης

Η βαθιά μάθηση απαιτεί μεγάλο όγκο επεξεργασίας δεδομένων όπως επίσης και ονοματοποιημένα αρχεία σε αντίθεση με την μηχανική μάθηση. Αυτό συμβαίνει γιατί χρησιμοποιώντας τα μοντέλα της βαθιάς μάθησης, τα οποία είναι πιο περίπλοκα, πρέπει να υπάρχει ένας ογκώδης χώρος αποθήκευσης χιλιάδων εικόνων έτσι ώστε να είναι επιτυχημένα και αληθή τα αποτελέσματα. Χρησιμοποιώντας λοιπόν μεγάλο όγκο δεδομένων πρέπει να υπάρχει ο κατάλληλος εξοπλισμός επεξεργασίας γραφικών. Κάρτες γραφικών τελευταίας τεχνολογίας είναι απαραίτητες, αφού μειώνουν αισθητά το χρόνο επεξεργασίας αυτών των δεδομένων.

Η μέθοδος επίλυσης των προβλημάτων με τη βοήθεια της μηχανικής μάθησης όμως δίνει την επιλογή εκμάθησης του μοντέλου σε περισσότερους ταξινομητές. Γίνεται επίσης γνωστή η εξαγωγή των συγκεκριμένων χαρακτηριστικών που θα χρησιμοποιηθούν για το καλύτερο αποτέλεσμα. Επιπλέον με τη μηχανική όραση δίνεται η δυνατότητα επιλογής διαφόρων συνδυασμών των προσεγγίσεων που

μπορούν να χρησιμοποιηθούν. Διαφορετικοί ταξινομητές και χαρακτηριστικά μπορούν επίσης να χρησιμοποιηθούν για να δουλέψουν συνδυαστικά με τα δεδομένα.

Συνοψίζοντας λοιπόν, με τη μηχανική μάθηση χρειάζονται λιγότερα δεδομένα απ' ό τι με τη βαθιά μάθηση και η εκμάθηση του μοντέλου γίνεται γρηγορότερα. Παράλληλα, με τη βαθιά μάθηση, που γίνεται όλο και πιο δημοφιλής, τα αποτελέσματα είναι ακριβέστερα και δε χρειάζεται να υπάρξει ανθρώπινη παρεμβολή ώστε να παρθούν αποφάσεις για τα κατάλληλα χαρακτηριστικά εκπροσώπησης του αντικειμένου, διότι μαθαίνονται αυτόματα από το μοντέλο. Έτσι, η ευθύνη της επιλογής των ταξινομητών είναι ελάχιστη σε σύγκριση με αυτής στη μηχανική μάθηση. Στη βαθιά μάθηση χρειαζόμαστε μεγαλύτερη ποσότητα δεδομένων, το οποίο σημαίνει περισσότερος χρόνος εκμάθησης του μοντέλου, όπως και καλύτερη ποιότητα υλικού. Γενικότερα, η επιλογή μεταξύ μηχανικής και βαθιάς μάθησης εξαρτάται από τα δεδομένα που υπάρχουν στην κατοχή του χρήστη, αλλά και το πρόβλημα το οποίο τίθεται να λύσει.

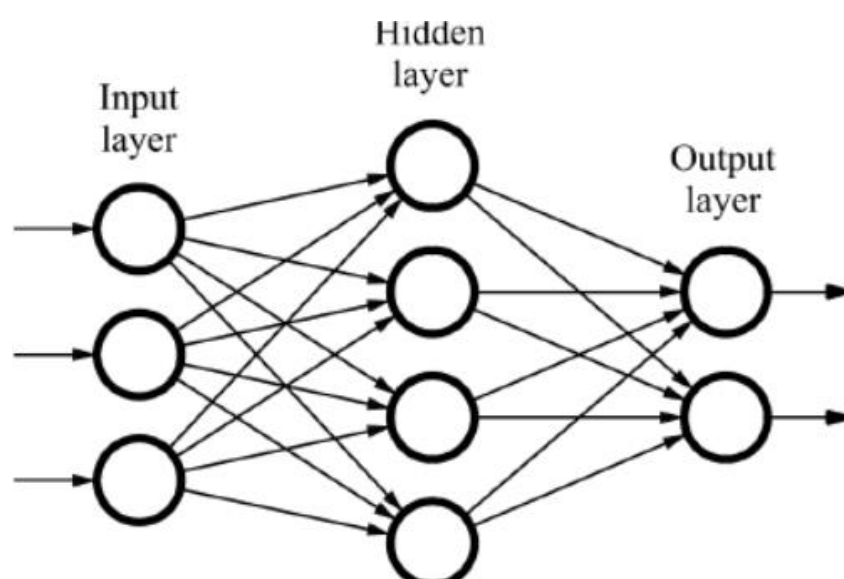
	Machine Learning	Deep Learning
Training dataset	Small	Large
Choose your own features	Yes	No
# of classifiers available	Many	Few
Training time	Short	Long

Εικόνα 32. Διαφορές μεταξύ μηχανικής και βαθιάς μάθησης

Νευρωνικά Δίκτυα

Στην σημερινή εποχή, κάποιο είδος νευρωνικού δικτύου αξιοποιείται, τουλάχιστον σε έναν βαθμό, σχεδόν καθημερινά από όλους μας. Εργασίες όπως αναγνώριση ομιλίας, μεταφράσεις κειμένου και πολλές άλλες, εκτελούνται σχεδόν αποκλειστικά από αυτές τις δομές.

Ένα νευρωνικό δίκτυο, συνήθως χαρακτηρίζεται από τρία διαφορετικά επίπεδα (Walczak & Cerga, 2003), το επίπεδο – layer εισόδου (input layer), το/τα κρυφό/κρυφά (hidden) και τέλος, την έξοδο/ (output layer) όπως φαίνεται και στην παρακάτω εικόνα:



Εικόνα 33. Τυπική δομή ενός Νευρωνικού Δικτύου

Το σημαντικότερο τους γνώρισμα εντοπίζεται στο ότι, λόγω του τρόπου λειτουργίας και εκπαίδευσής τους, έχουν την δυνατότητα να εξάγουν συμπεράσματα – βγάζουν νόημα/ αλληλοσυσχετισμούς μέσα από τεράστια ποσότητα πολύπλοκων και ασαφών δεδομένων, σε σύγκριση με τους αλγορίθμους μηχανικής μάθησης που είδαμε σε προηγούμενα κεφάλαια ή τον ανθρώπινο παράγοντα.

Εκπαίδευση ενός Νευρωνικού δικτύου

Κατά την έναρξη των εργασιών εκπαίδευσης ενός τέτοιου δικτύου, συνήθης τακτική αποτελεί η επιλογή των αρχικών βαρών. Τα ίδια, όπως θα δούμε παρακάτω, λαμβάνουν, αρχικά τυχαίες τιμές, αν και υπάρχει περίπτωση, ορισμένες τιμές βαρών σε συγκεκριμένα επίπεδα να παρθούν έτοιμες. Εδώ είναι σημαντικό, να αναλύσουμε

τις βασικότερες κατηγορίες μάθησης, οι οποίες έχουν αναφερθεί ήδη και στο κομμάτι της μηχανικής μάθησης, δηλαδή την προσέγγιση **εκπαίδευσης με επιτήρηση** και **χωρίς**. Συνήθως στην πρώτη κατηγορία, εντάσσονται, όπως είναι ήδη γνωστό, οι διαδικασίες εκείνες που αφορούν την ύπαρξη γνωστών επιθυμητών εξόδων και κατηγοριών. Αντίθετα η μη επιτηρούμενη μάθηση αφορά καταστάσεις όπου επαφίεται στο νευρωνικό δίκτυο να εξάγει συμπεράσματα σχετικά με τα δεδομένα εκπαίδευσης που του παρέχονται.

Εκπαίδευση με επιτήρηση

Στο πρώτο είδος εκπαίδευσης, πέραν από το σύνολο δεδομένων για εκπαίδευση που παρέχονται στην είσοδο, όπως προαναφέρθηκε, παρέχονται τόσο δεδομένα για την έξοδο. Η διαδικασία που λαμβάνει χώρα στη συνέχεια, αφορά την τυχαία επιλογή βαρών, την παραγωγή εξόδων και την σύγκρισή τους με τις αντίστοιχες επιθυμητές που έχουν δοθεί. Τα λάθη, δηλαδή οι διαφορές μεταξύ των πραγματικών και θεμιτών εξόδων, ‘μεταδίδονται’ όπως θα δούμε παρακάτω στις διαδικασίες του νευρωνικού δικτύου με αποτέλεσμα την διόρθωση των βαρών ώστε να ελαχιστοποιηθούν.

Εδώ, αρκετά σημαντικό είναι να αναφερθεί η μεγάλη ποσότητα και καλή ποιότητα δεδομένων που απαιτείται για την εκπαίδευση ενός νευρωνικού δικτύου στις περιπτώσεις εκπαίδευσης με επιτήρηση, καθώς σε αντίθεση περίπτωση ο ταξινομητής μας θα παρουσιάζει προβλήματα στην επίδοσή του.

Μη επιτηρούμενη ή προσαρμοστική εκπαίδευση

Η δεύτερη μεγάλη κατηγορία, αφορά τις περιπτώσεις όπου οι έξοδοι δεν είναι γνωστοί, και ονομάζεται μη επιτηρούμενη ή προσαρμοστική εκπαίδευση – μάθηση. Συγκεκριμένα, ενώ παρέχονται τα δεδομένα εισόδου, συνήθως λόγω της φύσης του προβλήματος δεν είναι δυνατόν να προσδιορίσουμε τον τύπο της εξόδου πόσο μάλλον να διακρίνουμε συγκεκριμένες κατηγορίες. Έτσι, βασιζόμαστε, ουσιαστικά στο δίκτυο να εξάγει συμπεράσματα σχετικά με τις ομαδοποιήσεις και αλληλοσυσχετίσεις, μια διαδικασία που ονομάζεται συχνά και ως αυτοοργάνωση ή προσαρμογή (για αυτό και η συγκεκριμένη εναλλακτική ονομασία σε αυτού του είδους την εκπαίδευση). Επί του παρόντος, χωρίς επίβλεψη η μάθηση δεν είναι καλά κατανοητή. Το βασικό χαρακτηριστικό αυτών των συστημάτων είναι ότι, αρκετές φορές τα συμπεράσματα που εξάγουν, μπορεί να μην γίνονται κατανοητά από τον

ανθρώπινο νου, με την έννοια ότι οι συσχετίσεις που παρουσιάζονται να μην βγάλουν κάποιο νόημα.

Εφαρμογή των ANN σε πραγματικό χρόνο

Πραγματοποιώντας μία σύντομη αναφορά στην πληθώρα των χρήσεων των νευρωνικών δικτύων θα παρατηρούσαμε ότι συναντώνται σε μία πληθώρα εφαρμογών. Στην περίπτωση της συγκεκριμένης διπλωματικής εργασίας, όμως, μας ενδιαφέρει κυρίως η χρήση τους σε πραγματικό χρόνο. Ονομαστικά μπορούμε να δούμε τις παρακάτω ευρύς κατηγορίες:

1. Χρήση σε προβλήματα παλινδρόμησης (Regression) είτε για την πρόβλεψη σε ζητήματα χρονοσειρών είτε καθαρά για λόγους μοντελοποίησης.
2. Διευκόλυνση πραγματοποίησης ορισμένων ενεργειών όπως απάντηση σε κλήσεις ή έλεγχος του συστήματος αναπαραγωγής μουσικής με συγκεκριμένες κινήσεις των χεριών. Διαχείριση τεχνολογιών όπως e book και πλοήγηση σε ιστοσελίδες με αντίστοιχες χειρονομίες
3. Επεξεργασία δεδομένων, ενέργειες φιλτραρίσματος, διαχωρισμού σήματος και ενεργειών συμπίεσης.
4. Συστήματα αναγνώρισης και ελέγχου, τόσο για την δική μας περίπτωση για την οποία γίνεται η μελέτη όσο και σε ευρύτερο πλαίσιο, όπως πχ στο σύστημα μεταφορών, καθώς βοηθούν στην υλοποίηση αυτο-ελεγχόμενων οχημάτων.
5. Σε παιχνίδια όπου η λήψη αποφάσεων διαδραματίζει σημαντικό ρόλο (τάβλι, σκάκι), όπου ήδη βαθιά νευρωνικά δίκτυα είναι σε θέση να νικήσουν τους καλύτερους επαγγελματίες.
6. Ιατρική βοήθεια τόσο όσον αφορά την πρόβλεψη ασθενειών όσο και κατά την διάγνωση ιατρικών εικόνων όπως πχ ακτινογραφίες
7. Οικονομικές εφαρμογές όπως εξόρυξη δεδομένων ή ανακάλυψη τυχών trends μεταξύ παραμέτρων, σε φρενήρεις ρυθμούς όπως αυτούς τους χρηματιστηρίου
8. Ακόμη, βοηθούν και στην περίπτωση όπου θέλουμε να αξιολογήσουμε μία συσκευή ως κόμβο πολυμέσων, δηλαδή να αναπαράγουμε το περιεχόμενο της σε διάφορες συσκευές. Εδώ, μπορούμε να το συνδυάσουμε και με την απομακρυσμένη διαχείριση που αναφέραμε πριν.
9. Σε διαδικτυακές υπηρεσίες που χρησιμοποιούμε καθημερινώς, όπως αλγόριθμοι στην πλατφόρμα YouTube, συνδυάζουν τέτοιες τεχνολογίες με

σκοπό να παρέχουν, σε πραγματικό χρόνο, το καλύτερο δυνατό προτεινόμενο βίντεο με βάση τον εκάστοτε χρήστη.

Πλεονεκτήματα Νευρωνικών Δικτύων

Ορισμένα από τα πλεονεκτήματα των νευρωνικών δικτύων που μπορούμε να ονομάσουμε είναι τα εξής:

- Λόγω της ικανότητας τους για προσαρμοσμένη μάθηση, είναι δυνατόν να μάθουν να εκτελούν καθήκοντα μόνο παρέχοντας εισόδους για δεδομένα προς εκπαίδευση, είτε μέσω, στο ίδιο πλαίσιο, της αρχικής εμπειρίας από τον ανθρώπινο παράγοντα.
- *Αυτο-οργάνωση*: Όπως προ είπαμε, ένα νευρωνικό δίκτυο μπορεί να δημιουργήσει τον δικό του τρόπο οργάνωσης ανεξαρτήτως της πολυπλοκότητας της εισόδου που του παρέχουμε και να παρέχει, κάποια εκπροσώπηση των πληροφοριών κατά την διάρκεια της εκπαίδευσης.
- Υπό τις κατάλληλες συνθήκες, είναι δυνατόν οι λειτουργίες του νευρωνικού δικτύου να γίνονται σε πραγματικό χρόνο, παρέχοντας έτσι ένα πολύ σημαντικό εργαλείο για μία πληθώρα εφαρμογών.
- Το βασικό χαρακτηριστικό αυτών των συστημάτων είναι ότι η εκπαίδευση δεν περιλαμβάνει διαδικασίες προγραμματισμού. Αυτό σημαίνει, ότι ακολουθώντας έναν αλγόριθμο βελτιστοποίησης και αλλάζοντας ορισμένες παραμέτρους, ο δημιουργός του συστήματος είναι σε θέση να παρατηρεί αλλαγές χωρίς να εμπλέκεται στο να προγραμματίσει κάποιο βήμα εξαρχής.
- Συνδυάζοντας τα παραπάνω, καταλαβαίνουμε ότι τα ίδια τα δίκτυα παρουσιάζουν, εφόσον χρειαστεί, μεγάλη ευελιξία στις πραγματικές συνθήκες ως προς την αλλαγή των παραμέτρων που τα εκπαιδεύουν, γεγονός που τα καθιστά κατάλληλα για την συνεχώς μεταβαλλόμενη πραγματικότητά μας.
- Σε περιπτώσεις όπου παραδοσιακές μέθοδοι όπως, ο προγραμματισμός και η χρήση της στατιστικής, αποτυγχάνουν να παρέχουν αποτελέσματα ή να εξάγουν συμπεράσματα, η φύση των παραπάνω μοντέλων είναι σε θέση να παρέχει, πιθανές λύσεις.
- Τέλος, σε συνδυασμό με το παραπάνω, αρκετές φορές οι λύσεις που παρέχονται από τις συγκεκριμένες διατάξεις, υπό κατάλληλες προϋποθέσεις φυσικά, είναι σε θέση να ξεπεράσουν, σε εκτιμητική ικανότητα τις παραδοσιακές μεθόδους.

Κατηγορίες Νευρωνικών δικτύων

Τα περισσότερα γνωστά, είδη νευρωνικών δικτύων είναι δύο

- Τα πολυεπίπεδα νευρωνικά δίκτυα perceptron
- Τα συνελκτικά νευρωνικά δίκτυα

Στη συνέχεια θα προχωρήσουμε σε ανάλυση της εκάστοτε δομής, με ενδελεχείς πληροφορίες για τις βασικές της λειτουργίες.

Συνελκτικά νευρωνικά δίκτυα

Στην περίπτωση των συνελκτικών νευρωνικών δικτύων, δηλαδή Convolutional Neural Networks (CNN), υπάρχουν σημαντικές ομοιότητες και διαφορές με την προηγούμενη κατηγορία και γενικότερα με τα παραδοσιακά τεχνητά νευρωνικά δίκτυα. Συγκεκριμένα, κάθε νευρώνας συνεχίζει να λαμβάνει μία είσοδο και να εκτελεί, αντίστοιχα μία λειτουργία όπως πριν ενώ ταυτόχρονα υπάρχει ακόμα η έννοια της συνάρτησης ενεργοποίησης στους νευρώνες που αυτο-βελτιστοποιούνται μέσω της μάθησης. Όπως και με τον αλγόριθμο του backpropagation κάθε νευρώνας εξακολουθεί να λαμβάνει μια είσοδο και να εκτελεί μια λειτουργία σε κάθε επίπεδο είτε με γραμμική είτε με μη γραμμική συνάρτηση. Στη συνέχεια, συνεχίζει να γίνεται η αυξομείωση του σκορ μέσω της γνωστής διαδικασίας ενώ στο τέλος πάλι θα υπάρχουν οι κλάσεις – κατηγορίες για το κάθε επίπεδο.

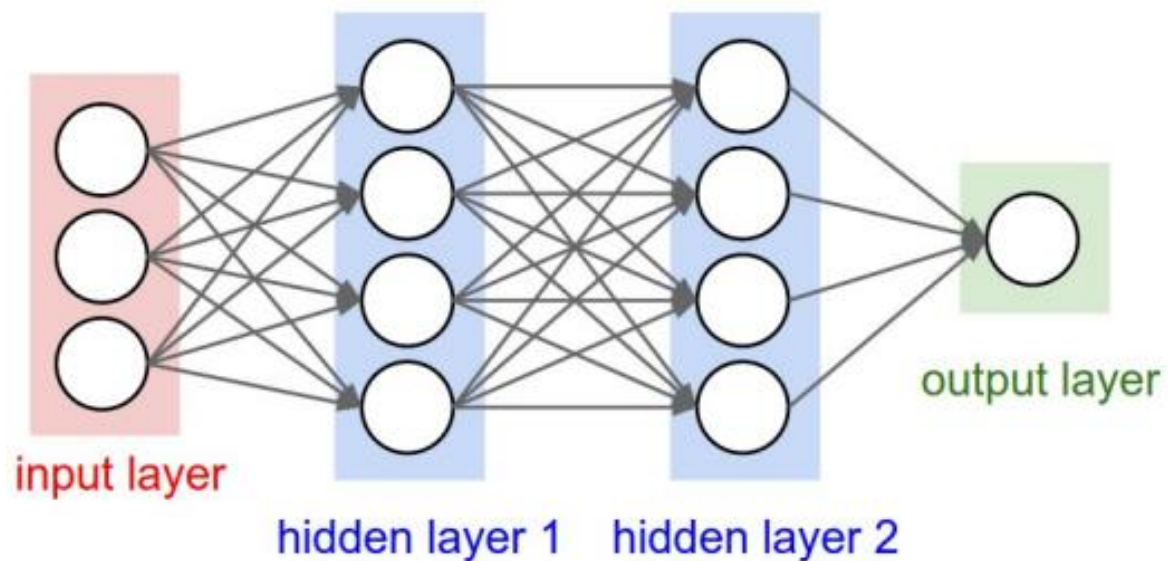
Η βασική διαφορά μεταξύ των αυτών κατηγοριών είναι ότι τα τελευταία, CNN χρησιμοποιούνται κυρίως στον τομέα της αναγνώρισης προτύπων μέσα σε εικόνες. Αυτό επιτρέπει την κωδικοποίηση ειδικών χαρακτηριστικών εικόνας τόσο στο κομμάτι της αρχιτεκτονικής καθιστώντας το δίκτυο ικανό εργασίες που εστιάζουν στην εικόνα - μειώνοντας παράλληλα τις παραμέτρους που απαιτείται για τη ρύθμιση του μοντέλου. Ένας από τους μεγαλύτερους περιορισμούς των παραδοσιακών μορφών του ANN είναι ότι τείνουν να ανταγωνίζονται με την υπολογιστική πολυπλοκότητα που απαιτείται για τον υπολογισμό των δεδομένων εικόνας, λόγω της πολύπλοκης δομής τους. Με δεδομένο ότι τα CNN εστιάζουν κυρίως στη βάση ότι θα περιληφθεί η εισαγωγή εικόνων, προσαρμόζουν αντίστοιχα και την αρχιτεκτονική τους ώστε να ταιριάζει καλύτερα την ανάγκη αντιμετώπισης του συγκεκριμένου τύπου δεδομένων.

Μία από τις βασικές διαφορές είναι ότι οι νευρώνες που στρώνονται μέσα στο CNN αποτελούνται από νευρώνες οργανωμένους σε τρεις διαστάσεις, τη χωρική διάσταση της εισόδου (ύψος και πλάτος) και το βάθος. Το βάθος δεν είναι ο συνολικός αριθμός επιπέδων εντός του ANN, αλλά στην τρίτη διάσταση του όγκου ενεργοποίησης. Σε αντίθεση με τα τυπικά ANNs, οι νευρώνες σε οποιοδήποτε δεδομένο στρώμα θα συνδεθεί μόνο σε μια μικρή περιοχή του στρώματος που προηγείται.

Πολυεπίπεδα Νευρωνικά δίκτυα Perceptron

Δομή νευρωνικού δικτύου Perceptron

Ο πολυεπίπεδος perceptron αποτελεί ένα αρκετά γνωστό και ιδιαίτερα χρήσιμο είδος νευρωνικού δικτύου (Gardner & RDorling, 1998). Η δομή του, συνήθως, ομοιάζει αυτή όπου, τα σήματα ή αλλιώς η πληροφορία μεταδίδονται εντός του δικτύου προς μία κατεύθυνση: από είσοδο στην έξοδο. Αυτή η αρχιτεκτονική ονομάζεται feedforward(εμπρόσθιας διάδοσης). Ένα χαρακτηριστικό σχήμα φαίνεται στην παρακάτω εικόνα.



Εικόνα 34. Πολυεπίπεδος Perceptron Εμπρόσθιας διάδοσης

Όπως φαίνεται στην εικόνα, μπορούμε να διακρίνουμε διαφορετικές κατηγορίες επιπέδων. Ονομαστικά:

1. Input Layer – Επίπεδο Εισόδου
2. Hidden Layer(s) – Κρυφά Επίπεδα
3. Output Layer – Επίπεδο Εξόδου

Αναλύοντας αυτές τις έννοιες, κατανοούμε ότι η πρώτη οντότητα, δηλαδή το επίπεδο εισόδου αφορά την εισαγωγή των αρχικών δεδομένων ως σειτ εκπαίδευσης στο δίκτυό μας. Τα επόμενα επίπεδα τα κρυφά, αφορούν της αριθμητικές πράξεις που πραγματοποιεί το δίκτυο, με τα κατάλληλα βάρη στον κάθε κόμβο. Εδώ, πρέπει να τονιστεί, αν και θα αναλυθεί στη συνέχεια ότι η τοποθέτηση όλο και περισσότερων κρυφών επιπέδων, αυξάνει σημαντικά την πολυπλοκότητα του νευρωνικού δικτύου. Τέλος, το επίπεδο εξόδου, αφορά τις κατηγορίες, στην περίπτωση της επιβλεπόμενης μάθησης, που πρόκειται να προβλεφθούν.

Σε αντίθεση με αυτά που μελετάμε τώρα, υπάρχουν και περιπτώσεις όπου ενδιάμεσα στα δίκτυα παρουσιάζονται δυνατότητες αποστολής των σημάτων/πληροφοριών και στις δύο κατευθύνσεις, μία μορφή που ονομάζεται και ανάδραση – feedback. Βέβαια, όπως είναι λογικό, η πολυπλοκότητα της αρχιτεκτονικής αυξάνεται σημαντικά στις συγκεκριμένες περιπτώσεις, ειδικά αν το συνδυάσουμε με περιπτώσεις πραγματοποίησεως υπολογισμών σε πραγματικό χρόνο.

Βέβαια, εδώ είναι σημαντικό να κατανοήσουμε το γιατί η συγκεκριμένη διάταξη είναι διάσημη. Η ισχύς του πολυεπίπεδου perceptron προκύπτει, κυρίως, από **μη γραμμικές** λειτουργίες ενεργοποίησης. Για αυτές έχουμε αναφερθεί προηγουμένως όποτε εδώ θα πραγματοποιηθεί απλά μία σύντομη αναφορά. Συγκεκριμένα, σχεδόν οποιαδήποτε μη γραμμική συνάρτηση μπορεί να χρησιμοποιηθεί για αυτόν τον σκοπό, εκτός από τις πολυωνυμικές συναρτήσεις, όπως αυτή της λογιστικής ή σιγμοειδούς, όπου καλό θα ήταν να τονίσουμε την ενέργεια των λειτουργιών αυτών για μικρές απόλυτες τιμές ως γραμμική. Αντίθετα, στις υψηλές απόλυτες τιμές, φέρουν τον ρόλο του κατωφλίου.

Αυτό δικαιολογεί την ιδιότητα του πολυεπίπεδου perceptron να ενεργεί ως μέθοδος καθολικής προσέγγισης. Το παραπάνω σε συνδυασμό με το θεώρημα Stone-Weierstrass (Cotter, 1990) που αφορά τα νευρωνικά δίκτυα και πώς μπορούν να υπολογιστούν ορισμένες πολυωνυμικές εκφράσεις μέσω αυτών, καθιστά τον πολυεπίπεδο perceptron αρκετά σημαντικό σαν εργαλείο για την επίλυση δύσκολων προβλημάτων.

Μία σημαντική έννοια που πρέπει να αναλύσουμε είναι πώς πραγματοποιείται όλη αυτή η διόρθωση των σφαλμάτων μεταξύ της επιθυμητής και της πραγματικής εισόδου, το λεγόμενο *backpropagation*

Ο αλγόριθμος backpropagation

Αναλύοντας συνοπτικά, μία διαδικασία κατηγοριοποίησης από ένα *feedforward* νευρωνικό δίκτυο, για χάριν ευκολίας θα αναφερθούμε αμιγώς στην εκπαίδευση με επιτήρηση, μπορούμε να διακρίνουμε τα εξής βήματα:

1. Αρχικοποίηση των τιμών των βαρών
2. Εκκίνηση νέας εποχής (epoch) εκπαίδευσης
3. Μεταφορά της πληροφορίας προς τα εμπρός
4. Προς τα πίσω διάδοση των λαθών
5. Προσαρμογή των βαρών

Στην συγκεκριμένη ενότητα, όπως είναι φανερό θα αναφερθούμε στα τελευταία 2 βήματα, δηλαδή το πώς πραγματοποιείται η διόρθωση της απόκλισης μεταξύ της θεμιτής και πραγματικής εισόδου, οδηγώντας στην κατασκευή ενός καλύτερου εκτιμητή. Για να γίνουν καλύτερα κατανοητές οι συγκεκριμένες διαδικασίες θα ήταν ωφέλιμο να χρησιμοποιηθεί ένα συγκεκριμένο παράδειγμα.

Τα δίκτυα μάθησης όσον αφορά τις δομές νευρωνικών δικτύων επιτυγχάνονται συνήθως με ένα εποπτευόμενο τρόπο που έχουμε ήδη αναφέρει. Μπορεί να θεωρηθεί ότι είναι διαθέσιμο ένα μαθησιακό περιβάλλον που περιέχει και τα δύο μοντέλα μάθησης και μοντέλα επιθυμητής παραγωγής αντιστοιχεί στην είσοδο (αυτό είναι γνωστό ως "στόχος" μοντέλα "). Έχουμε ήδη αναφέρει ότι η εκπαίδευση βασίζεται, η μάθηση βασίζεται συνήθως στην ελαχιστοποίηση των σφαλμάτων μέτρησης μεταξύ έξοδων δικτύου και επιθυμητών εξόδων. Αυτό συνεπάγεται, πίσω διάδοση μέσω ενός δικτύου παρόμοιου ή και ακριβώς ίδιου με αυτό που γίνεται η εκπαίδευση. Για αυτόν τον λόγο και το όνομα του αλγορίθμου είναι οπίσθιας διάδοσης - *backpropagation*.

Από ιστορική σκοπιά, η μέθοδος αυτή έχει προταθεί αρκετά παλιά, αλλά είχε αγνοηθεί σε σημαντικό βαθμό λόγω του μεγάλου υπολογιστικού κόστους, ενώ συνέχεια 'ανακαλύφθηκε' ξανά γύρω στα μέσα της δεκαετίας του 80 και ουσιαστικά

ξεκίνησε να αποτελεί ένα γενικά αποδεκτό εργαλείο για την εκπαίδευση του πολυεπίπεδο perceptron. Η ιδέα είναι να βρείτε το συνάρτηση ελάχιστου σφάλματος $e(w)$ σε σχέση με τα βάρη συνδέσεων.

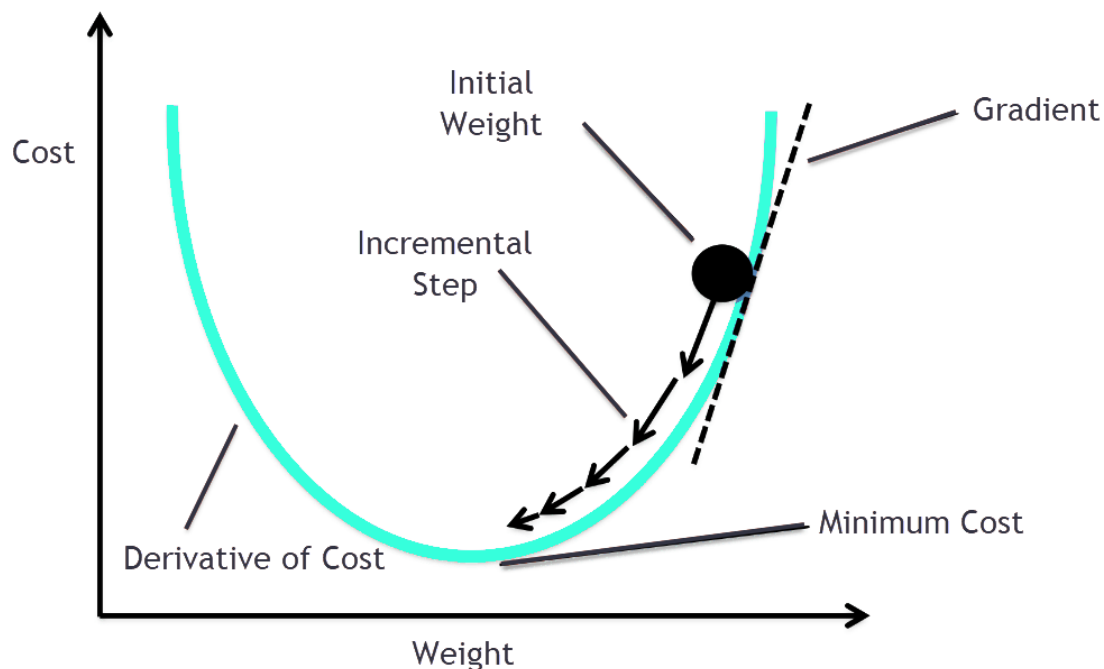
Μία σύντομη αναφορά για τα γενικά βήματα ενός πολυεπίπεδου perceptron πριν την ανάλυση του backpropagation είναι η παρακάτω:

Αρχικά, όλα τα βάρη του δικτύου, αρχικοποιούνται με τυχαίες τιμές, το εύρος των οποίων κατανέμεται ομοιόμορφα και σε μικρή έκταση. Στην περίπτωση όπου οι τιμές αυτές προσεγγίζουν το 0 οι αντίστοιχες κλίσεις που θα υπολογιστούν θα τείνουν και αυτές στο 0, στην περίπτωση όπου έχουμε κρυφά επίπεδα, γεγονός που θα οδηγήσει στην μη εκπαίδευση του δικτύου. Περισσότερο υποδεικνύονται προσπάθειες εκπαίδευσης, με διαφορετικά αρχικά βάρη, για να βρεθεί η καλύτερη τιμή για τη συνάρτηση κόστους (ελάχιστο σφάλμα). Αντίθετα, εάν οι αρχικές τιμές είναι μεγάλες, τότε αυτές τείνουν να κορεστούν με αποτέλεσμα, η σιγμοειδής συνάρτηση ενεργοποίησης που συνήθως χρησιμοποιείται να είναι πολύ μικρή. Λόγω του γεγονός ότι θα λειτουργήσει ως πολλαπλασιαστικός παράγοντας κατά τη διάρκεια της μαθησιακής διαδικασίας η παραπάνω τιμή, οι κορεσμένες μονάδες θα είναι σχεδόν μπλοκαρισμένες, η οποία καθιστά τη μάθηση πολύ αργή.

Σε μια νέα εποχή εκπαίδευσης, παρουσιάζει όλα τα παραδείγματα του σετ προπόνησης ο αλγόριθμος στο νευρωνικό δίκτυο. Στις περισσότερες περιπτώσεις, η εκπαίδευση του δικτύου περιλαμβάνει περισσότερες από μία εποχές. Κατά την εκπαίδευση, τα βάρη ρυθμίζονται μόνο μετά από όλα τα διανύσματα δοκιμής που θα εφαρμοστούν στο δίκτυο. Επομένως, για τις κλίσεις των βαρών, οι ίδιες πρέπει να απομνημονεύονται και να προσαρμόζονται μετά από κάθε μοντέλο στο αντίστοιχο σετ προπόνησης.

Σαν αποτέλεσμα τα βάρη θα αλλάξουν μόνο μία φορά (υπάρχει μια online παραλλαγή, πιο απλό, στο οποίο ενημερώνονται τα βάρη άμεσα, σε αυτήν την περίπτωση, τη σειρά με την οποία τα διανύσματα του δικτύου που παρουσιάζονται μπορεί να έχει σημασία. Όλες οι βαθμίδες των βαρών και το τρέχον σφάλμα αρχικοποιούνται με 0.

Εδώ είναι σημαντικό να αναφερθούμε στην έννοια της *Gradient Descent*. Η ίδια η έννοια είναι στενός συνδεδεμένη με τον αλγόριθμο backpropagation και την γενικότερη ιδέα της βελτιστοποίησης του συστήματος.



Εικόνα 35. Η έννοια της *Gradient Descent*

Ουσιαστικά αφορά την ελαχιστοποίηση του σφάλματος της παραγώγου που αναφέρθηκε παραπάνω. Για να την κατανοήσουμε καλύτερα, θα πρέπει να αναλύσουμε και ορισμένες έννοιες όπως minima και maxima (ολικά/τοπικά μέγιστα – ελάχιστα), συναρτήσεις σύγκλισης, συνθήκη παύσης και τον ρυθμό εκπαίδευσης που αναφέραμε.

Συνάρτηση Κόστους

Όπως έχουμε ήδη, δει η συνάρτηση κόστους αφορά τον βασικό τρόπο μέτρησης του σφάλματος μεταξύ της πραγματικής και επιθυμητής εξόδου. Προφανώς, κάθε αλγόριθμος βημάτων βελτιστοποίησης ενός νευρωνικού δικτύου, στοχεύει στην όσο το δυνατόν μεγαλύτερη μείωση της αριθμητικής τιμής αυτής της συνάρτησης, καθώς αυτό θα σημαίνει μικρότερη απόκλιση μεταξύ των δύο εξόδων και, εν τέλει, την παραγωγή εκτιμητών με ισχυρότερες δυνατότητες πρόβλεψης.

Στο παραπάνω παράδειγμα, αναφέραμε ήδη μία αρκετά διάσημη και ευρέως χρησιμοποιημένη συνάρτηση, αυτήν του μέσου τετραγωνικού κόστους (Mean Squared Error – MSE).

$$CF = \frac{1}{n} \sum_{i=1}^n (o_i - odes_i)^2$$

Στην περίπτωση του παραδείγματος βάλουμε για $n=2$. Συνήθως χρησιμοποιείται για προβλήματα που έχουν να κάνουν με παλινδρόμηση.

Σε παρόμοια λογική, χρησιμοποιείται και η λογαριθμική συνάρτηση κόστους ή εντροπίας:

$$CF = -\frac{1}{n} \sum_{i=1}^n p_i * \ln(pdes_i) + (1 - p_i) * \ln(1 - pdes_i)^2$$

Εδώ, η ίδια εντοπίζεται περισσότερο συχνά σε προβλήματα ταξινόμησης. Ως κλίση – Gradient ονομάζουμε το διάνυσμα, για n μεταβλητές ενός προβλήματος, το διάνυσμα μήκους n που αναπαριστά την διεύθυνση σύμφωνα με την οποία το κόστος της συνάρτησης αυξάνεται πιο γρήγορα.

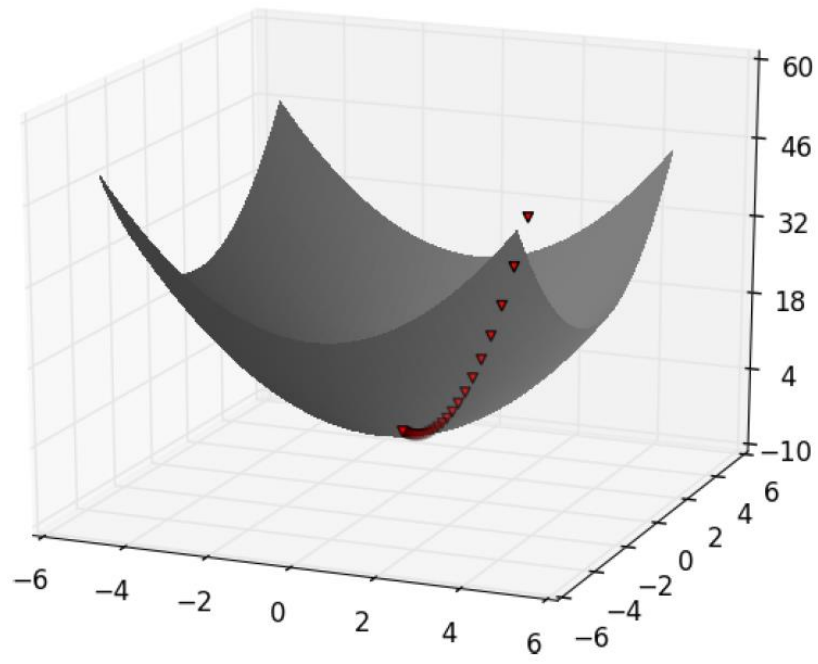
Για παράδειγμα, έστω μία γραμμική εξίσωση:

$$y = ax + b, \text{ με τιμές } a \text{ και } b \text{ γνωστές}$$

Και έστω ότι θέλουμε σε αυτήν να βρούμε το ελάχιστο σφάλμα οπότε η παραπάνω είναι και η συνάρτηση κόστους. Στην συγκεκριμένη περίπτωση η κλίση θα είναι προφανώς η:

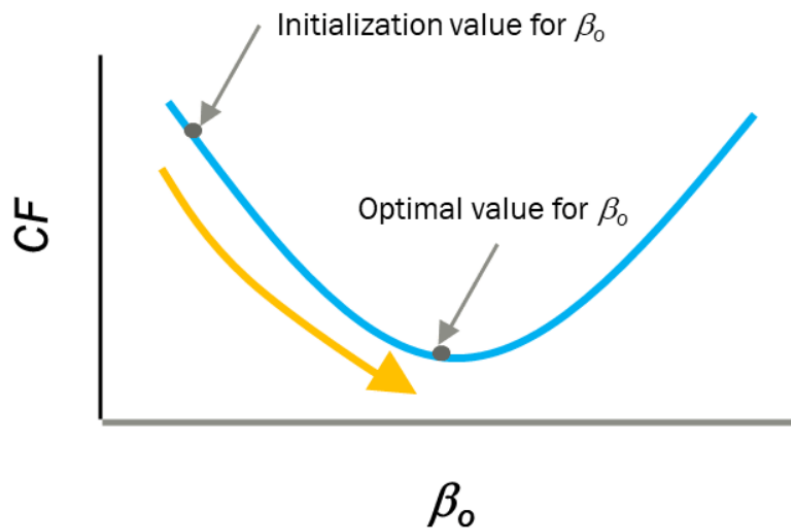
$$\frac{dy}{dx} = a$$

Στην περίπτωση αυτή, παίρνουμε ένα, αρχικό τυχαίο σημείο και κοιτάμε να δούμε πώς μπορούμε να φτάσουμε σε κάποιο ελάχιστο, ελπίζοντας ότι θα είναι ολικό.



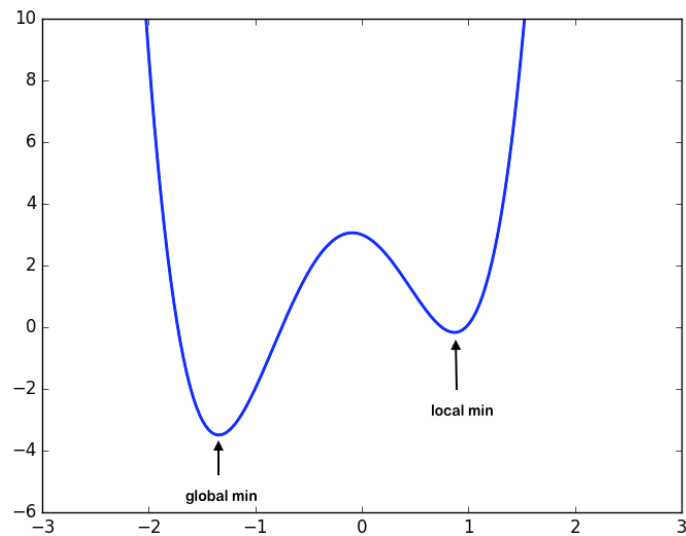
Εικόνα 36. Κίνηση *Gradient Descent* σε τρισδιάστατο πρόβλημα

Οι αλλαγές των βαρών-biases είναι αυτές που σχετίζονται με τα τοπικά και ολικά ελάχιστα



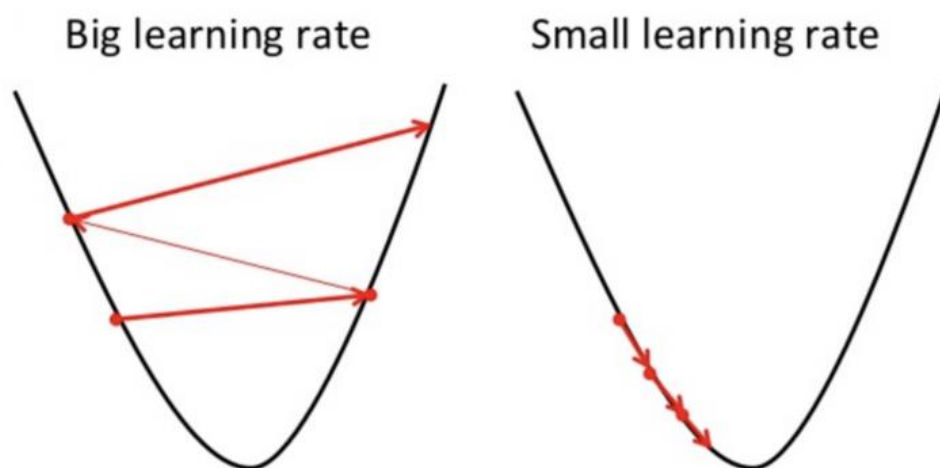
Εικόνα 37. Ολικό ελάχιστο – Βέλτιστη τιμή bias

Συγκεκριμένα, για τα τοπικά ελάχιστα, έστω μία συνάρτηση $g(x)$ η οποία μπορεί να διαθέτει ένα τοπικό και ένα ολικό μέγιστο ή ελάχιστο όπως φαίνεται στην εικόνα 38. Μπορεί να γίνει κατανοητό, από το παρακάτω σχήμα, ότι υπάρχει περίπτωση μία συνάρτηση κόστους, να 'πέσει' σε ένα τοπικό ελάχιστο και ο ταξινομητής να θεωρεί ότι αυτή είναι η βέλτιστη λύση. Για τον λόγο αυτό είναι ιδιαίτερα σημαντικές έννοιες όπως ο ρυθμός εκπαίδευσης και τρόποι βελτιστοποίησης αυτού.



Εικόνα 38. Ολικά/Τοπικά Ελάχιστα

Η αριθμητική τιμή του ρυθμού εκπαίδευσης, πρέπει να τίθεται σε ένα όριο, όχι ιδιαίτερα ψηλά, γιατί υπάρχει μεγάλη πιθανότητα λόγω των μεγάλων βημάτων που λαμβάνουν χώρα να μην εντοπιστεί ποτέ το τοπικό ελάχιστο. Αντίθετα στην περίπτωση πολύ μικρών βημάτων, υπάρχει μεγαλύτερη πιθανότητα να βρεθεί ένα ελάχιστο αν και εδώ ελλοχεύουν κίνδυνοι που κυρίως σχετίζονται με την αδυναμία εντοπισμού του ολικού ελαχίστου.



Εικόνα 39. Η διαφορά μεταξύ ακραίων τιμών ρυθμού εκπαίδευσης

Στις περιπτώσεις όπου έχουμε φτάσει σε ένα τοπικό ή ολικό ελάχιστο, ή τουλάχιστον σε ένα εύρος τιμών αρκετά κοντά σε αυτό τότε, συχνά λέμε ότι η συνάρτηση κόστους

έχει συγκλίνει. Κατά την σύγκλιση, ουσιαστικά, πλέον, πραγματοποιείται σπατάλη των υπολογιστικών πόρων των μηχανημάτων υπό το πρίσμα της μη σημαντικής βελτίωσης των αποτελεσμάτων.

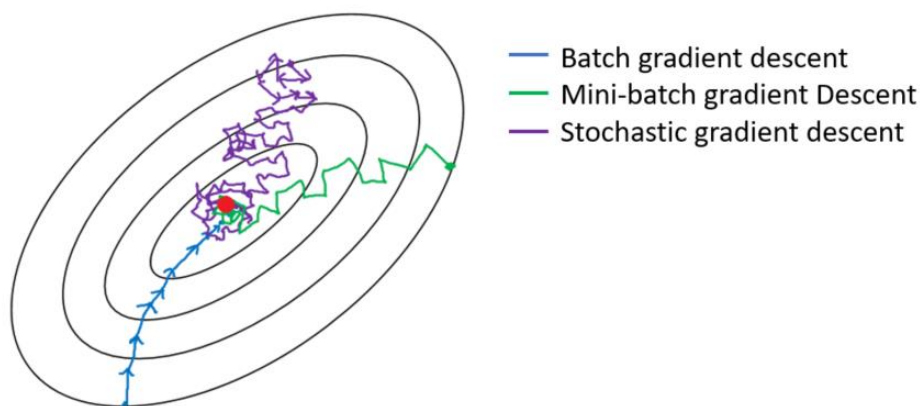
Για την Gradient Descent, υπάρχουν ορισμένες βασικές τακτικές όπως:

- *Batch Gradient Descent*
- *Stochastic Gradient Descent*
- *Mini – batch Gradient Descent*

Η πρώτη, ενώ ομοιάζει στον κανονικό απλό αλγόριθμο, αφού υπολογίζει το σφάλμα σε κάθε παράδειγμα του σετ εκπαίδευσης, ενημερώνει το μοντέλο μόνο μετά το πέρας της ολοκλήρωσης ενός κύκλου της εκπαίδευσης (training epoch – batch). Αυτό, ενώ παρέχει τα θετικά ότι είναι αρκετά οικονομική από θέμα υπολογιστικού κόστους, μπορεί να οδηγήσει σε πρόωρη σύγκλιση των αλγόριθμο σε μη βέλτιστη λύση σε σύγκριση με τις κλασικές μεθόδους

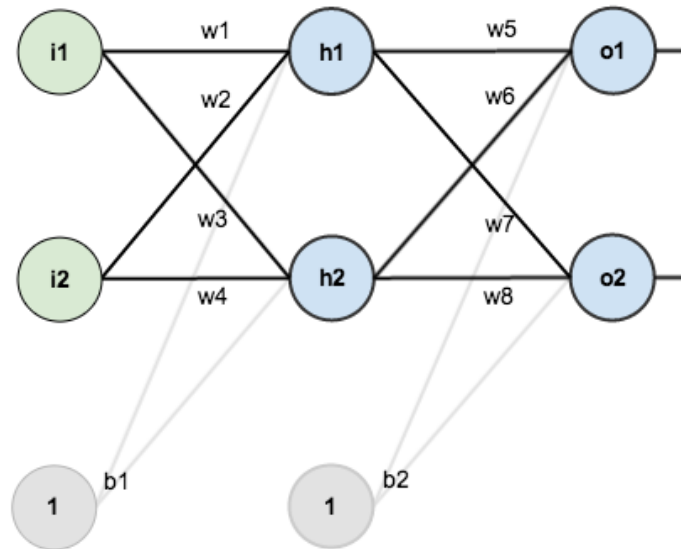
Από την άλλη, η Stochastic Gradient Descent ανανεώνει τα δεδομένα για κάθε παράδειγμα του σετ, γεγονός που σημαίνει ότι ενώ σίγουρα είναι περισσότερο απαιτητική από θέμα κόστους, έχει μεγαλύτερες πιθανότητες να συγκλίνει σε κάποιο ελάχιστο συντομότερα.

Τέλος η mini Batch GD που ουσιαστικά χωρίζει το σετ σε μικρότερα κομμάτια – batches τα οποία αξιοποιούνται για να υπολογίζονται τόσο το σφάλμα όσο και να ενημερώνουν το μοντέλο.



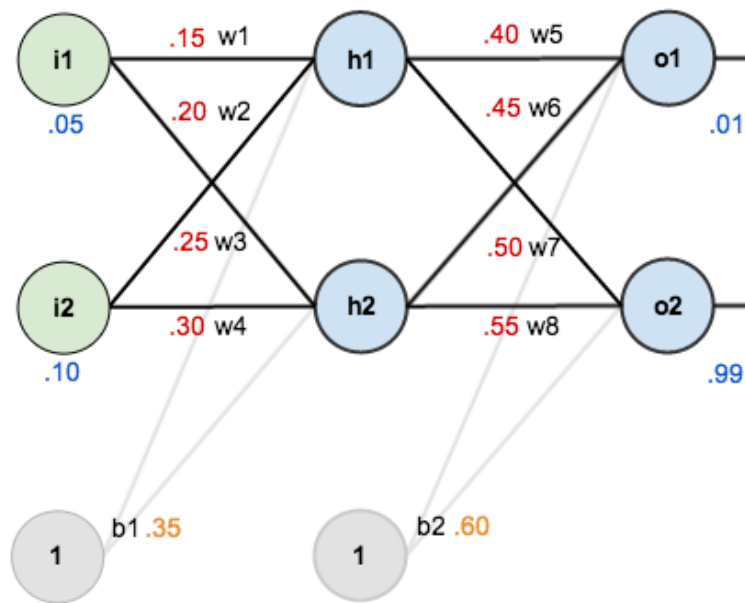
Εικόνα 40. Αναπαράσταση των διαφορετικών τύπων GD

Έστω, λοιπόν το παρακάτω νευρωνικό δίκτυο με ένα επίπεδο εισόδου, εξόδου και κρυφό.



Εικόνα 41. Τυχαίο FeedForward Νευρωνικό Δίκτυο

Μίας και αφορά άλλη ενότητα, έστω και ότι πραγματοποιείται η αρχική τοποθέτηση, τυχαίων τιμών για τα βάρη του νευρωνικού δικτύου $w1, w2, w3, w4, w5, w6, w7, w8$ και αντίστοιχα υπολογίζονται, κάπως οι τιμές των biases, όπως φαίνεται στην παρακάτω εικόνα:



Εικόνα 42. Αρχικοποίηση βαρών και υπολογισμός των biases

Έστω ακόμη, όπως φαίνεται και στην εικόνα 36 ότι οι τιμές που μας παρέχονται ως είσοδοι είναι οι 0.05 και 0.1 και οι **επιθυμητές** έξοδοι που θέλουμε είναι 0.01 στο o1 και 0,99 στο o2 αντίστοιχα. Ακόμη, έστω ότι για συνάρτηση ενεργοποίησης χρησιμοποιούμε την λογιστική συνάρτηση που έχουμε πει.

Υπολογίζουμε αρχικά την τιμή του κάθε νευρώνα του κρυφού δικτύου, δηλαδή:

$$h_1 = w_1 * i_1 + w_2 * i_2 + b_1 * 1$$

Και

$$h_2 = w_3 * i_1 + w_4 * i_2 + b_1 * 1$$

Υπολογίζουμε τις αντίστοιχες τιμές και έπειτα την κάθε τιμή την περνάμε από την λογιστική συνάρτηση που προ είπαμε πριν, δηλαδή, έστω h η τιμή τότε το out του θα είναι:

$$out_h = \frac{1}{1 + e^{-h}}$$

Στη συνέχεια αφού έχουμε υπολογίσει τα out για τις τιμές των δύο αυτών νευρώνων, εκτελούμε ακριβώς την ίδια διαδικασία, φυσικά χρησιμοποιώντας άλλους παράγοντες, για τους νευρώνες της εξόδου:

$$o_1 = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

Και

$$o_2 = w_7 * out_{h1} + w_8 * out_{h2} + b_2 * 1$$

Αποκτώντας έτσι τα τελικά αποτελέσματα.

Στη συνέχεια, χρησιμοποιώντας μία συνάρτηση κόστους, εδώ έστω αυτή του **τετραγώνου**, αθροίζουμε το λάθος για κάθε νευρώνα του επιπέδου εξόδου (πολλαπλασιάζουμε με $\frac{1}{2}$ λόγω παραγωγίσιμης):

$$E_{ολικό} = \sum \frac{1}{2} (\xi\sigma\delta\sigma_{\epsilon\pi\iota\theta\upsilon\mu\eta\tau\acute{\eta}} - \xi\sigma\delta\sigma_{\pi\rho\alpha\gamma\mu\alpha\tau\iota\kappa\acute{\eta}})^2$$

Έχοντας, πλέον το ολικό σφάλμα, τώρα στοχεύουμε στο να διορθώσουμε τα βάρη στο δίκτυο, ώστε να το μειώσουμε. Εδώ θα πρέπει να χρησιμοποιήσουμε έννοιες όπως η *μερική παράγωγος* και ο *κανόνας αλυσίδας*.

Η μερική παράγωγος μίας συνάρτησης πολλαπλών παραμέτρων, είναι μία παράγωγος όπου κατά την διάρκεια της παραγωγίσιμης, λαμβάνουμε και θεωρούμε μία από τις παραμέτρους εισόδου ως μεταβλητή και τις υπόλοιπες τις θεωρούμε σταθερές. Για παράδειγμα, έστω μία συνάρτηση f

$$f(x, y) = x^2 + y^2$$

Η μερική παράγωγος ως προς x θα είναι $2x$ ενώ η μερική παράγωγος ως προς y , $2y$.

Αντίστοιχα, ο κανόνας αλυσίδας αποτελεί μία μέθοδο ώστε να βρούμε την ολική παράγωγο μίας σύνθετης συνάρτησης. Με τον όρο σύνθετη ονομάζουμε μία συνάρτηση η οποία μπορεί να γραφτεί ως συνδυασμός άλλων συναρτήσεων. Για παράδειγμα, έστω ότι έχουμε την συνάρτηση:

$$h(x) = \cos(x^2) = f(g(x))$$

Αν θεωρήσουμε την $f(x)$ ως:

$$f(x) = \cos(x)$$

Και την $g(x)$ ως:

$$g(x) = x^2$$

Τότε βλέπουμε ότι ισχύει:

$$\cos(x^2) = f(g(x))$$

Έτσι, για την προηγούμενη υποθετική συνάρτηση:

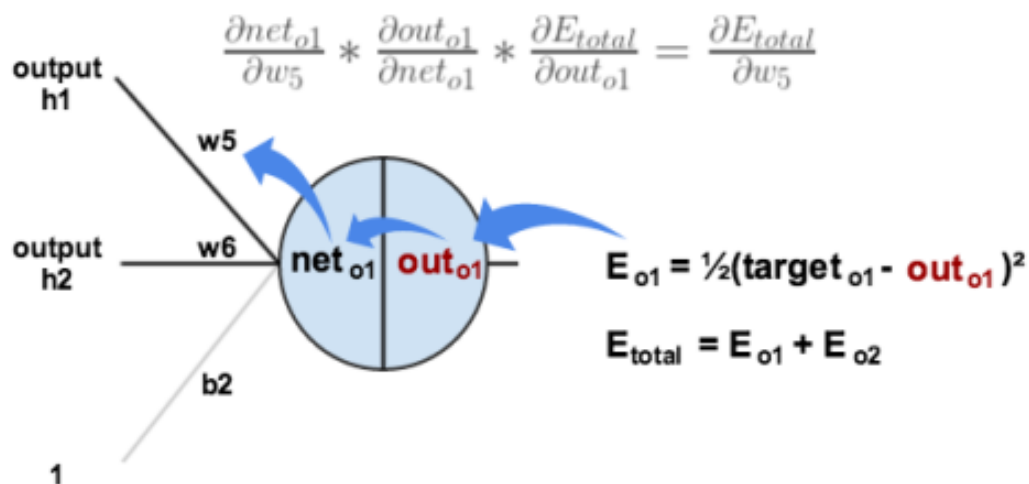
$$h(x) = f(g(x))$$

Σύμφωνα με τον κανόνα αλυσίδας, η παράγωγος ως προς x θα είναι:

$$f'(g(x)) * g'(x)$$

Επιστρέφοντας στο παράδειγμά μας, είμαστε, ακόμη στο επίπεδο εξόδου και θέλουμε να δούμε πώς οι αλλαγές σε ένα από τα βάρη w_5, w_6 για το o_1 θα επηρεάσουν το ολικό σφάλμα.

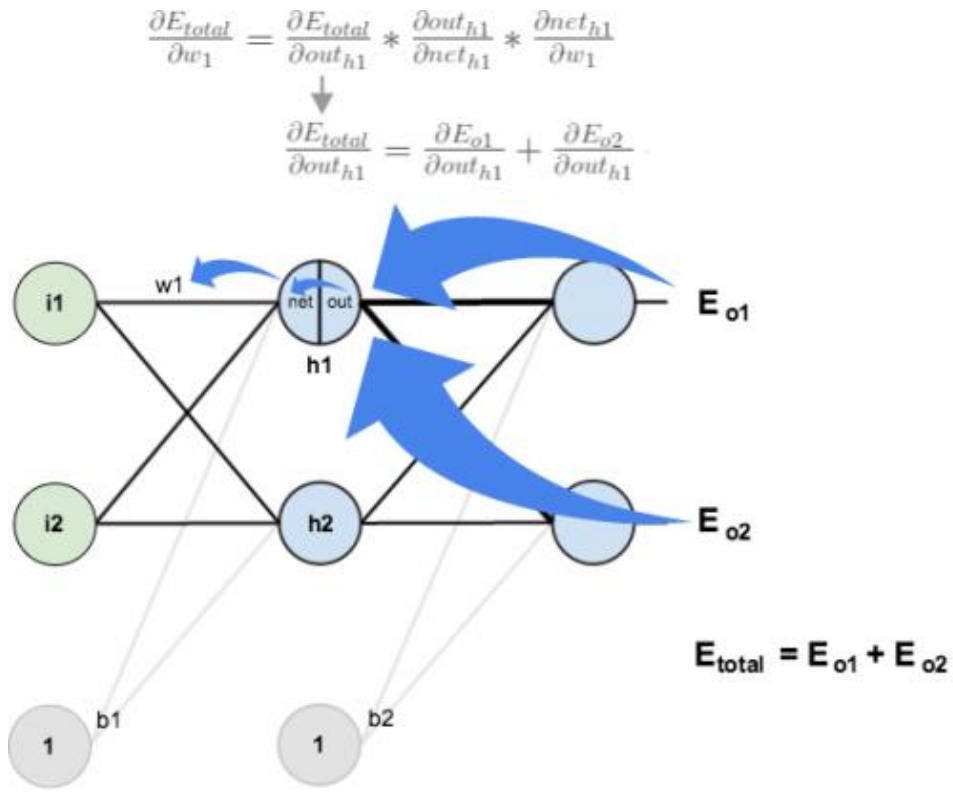
$$\frac{\partial E_{ολικό}}{\partial w_6} = \frac{\partial E_{ολικό}}{\partial out_{o1}} * \frac{\partial out_{o1}}{\partial o1} * \frac{\partial o1}{\partial w_6}$$



Εικόνα 43. Αναπαράσταση του πρώτου βήματος της μετάδοσης του σφάλματος προς τα πίσω

Αφού με αυτόν τον τρόπο υπολογίσουμε το σφάλμα για τον έναν νευρώνα, πολλαπλασιάζουμε την τελική του τιμή με μία παράμετρο που λέγεται **ρυθμός μάθησης**, ενώ η ίδια διαδικασία αξιοποιείται και για τα άλλα βάρη.

Η συνέχεια του backpropagation αλγορίθμου, πραγματοποιείται για το κρυφό επίπεδο, όπως φαίνεται παρακάτω:



Εικόνα 44. Μετάδοση του σφάλματος στο κρυφό επίπεδο

Η μέθοδος για τα βάρη w_1, w_2, \dots είναι ίδια, με την διαφορά ότι πλέον, εδώ, η έξοδος κάθε κρυφού νευρώνα επηρέαζε τις αντίστοιχες όλων των αντίστοιχων του επιπέδου εξόδου. Οπότε για τον νευρώνα h_2 , για παράδειγμα:

$$\frac{\partial E_{ολικό}}{\partial out_{h2}} = \frac{\partial E_{o1}}{\partial out_{h2}} + \frac{\partial E_{o2}}{\partial out_{h2}}$$

Όπου:

$$\frac{\partial E_{o2}}{\partial out_{h2}} = \frac{\partial E_{o2}}{\partial h_2} * \frac{\partial h_2}{\partial out_{h2}}$$

Και αντίστοιχα για το E_{o1} . Έτσι είμαστε σε θέση να υπολογίσουμε το συνολικό κόστος, για το βάρος ας πούμε w_3 :

$$\frac{\partial E_{total}}{\partial w_2} = \frac{\partial E_{total}}{\partial out_{h2}} * \frac{\partial out_{h2}}{\partial h_2} * \frac{\partial h_2}{\partial w_2}$$

Ενημερώνουμε το w2 πολλαπλασιάζοντάς το και αυτό με τον αντίστοιχο ρυθμό εκπαίδευσης και πραγματοποιούμε την ίδια διαδικασία για όλα τα υπόλοιπα βάρη.

Αναγνώριση προσώπου

Χαρακτηριστικά για αναγνώριση προσώπου

Για την διαδικασία αναγνώρισης προσώπων και αντικειμένων απαιτούνται κάποιες προϋποθέσεις. Μία απ' αυτές τις προϋποθέσεις είναι να βρεθούν συγκεκριμένα, αξιόπιστα, ανιχνεύσιμα κοινά χαρακτηριστικά ανάμεσα σε στις εικόνες που υπόκεινται σε επεξεργασία.

Τα χαρακτηριστικά σημεία χρησιμοποιούνται για κάθε είδους λειτουργία στην «όραση» του υπολογιστή, όπως ευθυγράμμιση της εικόνας, ανακατασκευή 3D, ανίχνευση κίνησης, αναγνώριση αντικειμένων στο χώρο, ευρετήρια βάσης δεδομένων και την πλοήγηση του ρομπότ. Υποθέτοντας πως υπάρχουν δύο εικόνες, βασική επιδίωξη είναι η κατανόηση του μετασχηματισμού που υπάρχει από τη μία εικόνα στην άλλη. Για να γίνει εφικτό αυτό πρέπει να υπάρχουν αντίστοιχα ίδια σημεία και στις δύο εικόνες. Ο τρόπος για να βρεθούν αυτά τα σημεία είναι η χρήση τοπικών χαρακτηριστικών. Ο σκοπός είναι να προσδιοριστούν και να βρεθούν, με ακρίβεια και αξιοπιστία, τα σημεία σε μια εικόνα που μπορούν να βρεθούν και σε άλλες εικόνες. Η χρήση των τοπικών αυτών χαρακτηριστικών είναι αναγκαία για τον υπολογισμό της θεμελιώδης μήτρας του υπολογιστή, στην διεκπεραίωση των σημείων και το που έχουν βρεθεί ανάμεσα στις εικόνες, για τον υπολογισμό της κίνησης της κάμερας όσον αφορά το βάθος και τα κινούμενα αντικείμενα στο χώρο, όπως επίσης και για την κατασκευή μιας πανοραμικής εικόνας.

Για τη διαδικασία κατασκευής μιας πανοραμικής εικόνας, πολλές εικόνες μαζί ευθυγραμμίζονται και επικολλώνται η μία στην άλλη. Η βασική διαδικασία είναι να βρεθούν για αρχή κάποια χαρακτηριστικά ή αλλιώς σημεία ενδιαφέροντος. Εντοπίζοντας αυτά τα σημεία θα πρέπει να βρεθούν αντίστοιχα και στις υπόλοιπες εικόνες τα οποία θα λειτουργήσουν ως ζεύγη, έτσι ώστε να γίνει κατανοητή η μετατροπή της μιας εικόνας στην άλλη. Χρησιμοποιώντας αυτά τα ζεύγη γίνεται ευθυγράμμιση ανάμεσα στις εικόνες. Παρ' όλα αυτά υπάρχουν δύο προβλήματα που κάνουν αυτή την αντιστοίχιση των χαρακτηριστικών δύσκολη. Δεν είναι εφικτή η αντιστοίχιση αν δεν μπορούν να βρεθούν ίδια σημεία ενδιαφέροντος ανάμεσα στις εικόνες. Γι' αυτό λοιπόν χρειάζεται ένας επαναλαμβανόμενος ανιχνευτής, ο οποίος θα

εντοπίζει τα σημεία ενδιαφέροντος που θα είναι όμοια στις εικόνες. Το δεύτερο πρόβλημα που παρουσιάζεται είναι η σωστή αντιστοίχιση των σημείων ενδιαφέροντος με αυτά που έχουν βρεθεί στις υπόλοιπες εικόνες. Ο τρόπος για να λυθεί αυτό είναι με τη χρήση ενός περιγραφέα. Ο περιγραφέας θα πρέπει να είναι διακριτός και να περιγράφει ακριβώς το σημείο ή αλλιώς το χαρακτηριστικό της εικόνας, αλλά και αξιόπιστος.

Ένα καλό και σωστά αποδεδειγμένο χαρακτηριστικό σημείο έχει συγκεκριμένες ιδιότητες που το διακρίνουν. Ως πρώτη ιδιότητα είναι η επαναληψιμότητα και ακρίβεια, όπου το ίδιο χαρακτηριστικό σημείο πρέπει να βρεθεί σε πολλαπλές εικόνες της ίδιας σκηνής παρά τις γεωμετρικές και φωτομετρικές αλλαγές που υπάρχουν. Επόμενη ιδιότητα που θα πρέπει να έχει είναι το κατά πόσο εξέχον είναι το χαρακτηριστικό έτσι ώστε η περιγραφή του να είναι μοναδική, με άλλα λόγια το πόσο διακριτός θα πρέπει να γίνει στον περιγραφέα. Κάθε χαρακτηριστικό θα πρέπει να μπορεί να περιγράφει μοναδικά για να είναι εύκολο στην αναγνώριση του. Επίσης, τα χαρακτηριστικά θα πρέπει να είναι γενικά πολύ λιγότερα σε σύγκριση με τον αριθμό των pixel σε μια εικόνα, ώστε να τα κάνουν αποτελεσματικά. Τέλος, θα πρέπει το χαρακτηριστικό σημείο να καταλαμβάνει ένα σχετικά μικρό μέρος της εικόνα, για να γίνεται η περιγραφή του σημείου αυτό όσο πιο απλά γίνεται.

Ταξινομητές για αναγνώριση προσώπου

Τι είναι ταξινόμηση

Η αναγνώριση διακρίνεται ως ένα πρόβλημα ταξινόμησης (classification). Παρατηρώντας την επίλυση των προβλημάτων στη μηχανική μάθηση αλλά και στη στατιστική, διακρίνεται το πρόβλημα του προσδιορισμού σε ποιο σύνολο κατηγοριών κατατάσσεται μια νέα παρατήρηση. Αυτό το πρόβλημα είναι η ταξινόμηση. Έχοντας ως αρχική βάση ένα σετ εκπαίδευσης των δεδομένων που περιέχει παρατηρήσεις ή περιπτώσεις των οποίων η κατηγορία μέλους είναι γνωστή, το μοντέλο θα πρέπει να αναγνωρίσει τη νέα παρατήρηση και να τη κατατάξει στην κατάλληλη κατηγορία. Η ταξινόμηση είναι ένα παράδειγμα αναγνώρισης μοτίβων, όπως για παράδειγμα η ταξινόμηση ηλεκτρονικού ταχυδρομείου σε εισερχόμενα και ανεπιθύμητα μηνύματα.

Στη μηχανική μάθηση συγκεκριμένα, οι παρατηρήσεις είναι συχνά γνωστές ως περιπτώσεις, οι ερμηνευτικές μεταβλητές ονομάζονται χαρακτηριστικά (ομαδοποιούνται σε ένα χαρακτηριστικό διάνυσμα), και οι πιθανές κατηγορίες που είναι να προβλεφθούν ονομάζονται κλάσεις. Η ταξινόμηση θεωρείται ένα παράδειγμα εποπτευόμενης μάθησης (supervised learning). Όπως αναφέρθηκε παραπάνω, ένα σύνολο δεδομένων σωστά προσδιορισμένων παρατηρήσεων είναι διαθέσιμο για την εκπαίδευση ενός αλγοριθμικού μοντέλου. Αυτή η τεχνική ταξινόμησης ονομάζεται εποπτευόμενη ταξινόμηση (supervised classification). Το κατά πόσο «σωστά» προσδιορισμένο είναι το σύνολο δεδομένων των παρατηρήσεων εξαρτάται από δύο περιπτώσεις. Αρχικά, από το είδος των λαθών που κάνει ο ταξινομητής, και έπειτα, από το κόστος των λαθών που επέρχεται με αυτά. Γνωρίζοντας λοιπόν τα επιθυμητά σεντ εκπαίδευσης, τίθεται αναγκαία η ελαχιστοποίηση της λάθος ταξινόμησης. Τα λάθη ταξινόμησης πρέπει να γίνονται σπανιότερα, αλλά και όσο μεγαλύτερο αντίκτυπο έχει αυτό το λάθος τόσο πιο σπάνια θα πρέπει να γίνεται. Αντίστοιχα, η μη επιβλεπόμενη διαδικασία ονομάζεται ομαδοποίηση, και περιλαμβάνει ομαδοποίηση των δεδομένων σε κατηγορίες με βάση κάποιο μέτρο της εγγενούς ομοιότητας ή της απόστασης.

Για να διαχειριστούμε όσο το δυνατόν γίνεται καλύτερα την ταξινόμηση, τη διαχωρίζουμε σε δύο μεθόδους. Η πρώτη μέθοδος ονομάζεται παραγωγική. Εκεί χρησιμοποιούνται τα δεδομένα εκπαίδευσης των διαφορετικών κλάσεων ώστε να δημιουργηθούν τα αντίστοιχα μοντέλα. Για παράδειγμα υπάρχουν οι κλάσεις A, B, Γ, Δ και αντίστοιχα έχουμε και τα μοντέλα A, B, Γ, Δ που δημιουργούνται από τα δεδομένα της κάθε κλάσης ξεχωριστά. Έπειτα, αφού έχει δοθεί ένα νέο παράδειγμα, συγκρίνεται με αυτά τα μοντέλα που έχουν δημιουργηθεί και οποιοδήποτε από αυτά το εξηγεί καλύτερα επιλέγεται ως το κατάλληλο. Ονομάζεται έτσι κυρίως γιατί περιγράφει όλες τις πιθανές περιπτώσεις που μπορεί να εμπεριέχονται στην κάθε κλάση. Η άλλη μέθοδος που χρησιμοποιείται είναι η διακριτική. Εκεί διακρίνονται τα δεδομένα μιας κλάσης με τα αυτά των υπόλοιπων κλάσεων, ώστε να κατασκευαστεί ένα επιθυμητό όριο μεταξύ των μοντέλων. Για παράδειγμα έχουμε την κλάση A και τις υπόλοιπες που δεν ανήκουν σ' αυτήν. Πρέπει να διακριθεί η διαφορά μεταξύ της κλάσης A με τις υπόλοιπες και να γίνει γνωστός ο τρόπος με τον οποίο διακρίνεται αυτή η διαφορά. Έτσι δεν αφιερώνεται χρόνος στη μοντελοποίηση της κάθε κλάσης

ξεχωριστά αλλά το πως να τις διακρίνει. Εξού και η ονομασία της, γιατί εστιάζει στην εκμάθηση του ορίου μεταξύ των κατηγοριών.

Βασικές κατηγορίες ταξινομητών

Η αναγνώριση προσώπου μπορεί να πραγματοποιηθεί με μία ποικιλία διαφορετικών μεθόδων. Η πρώτη που θα δούμε είναι το Holistic ταίριασμα (matching).

Holistic Matching (appearance-based method)

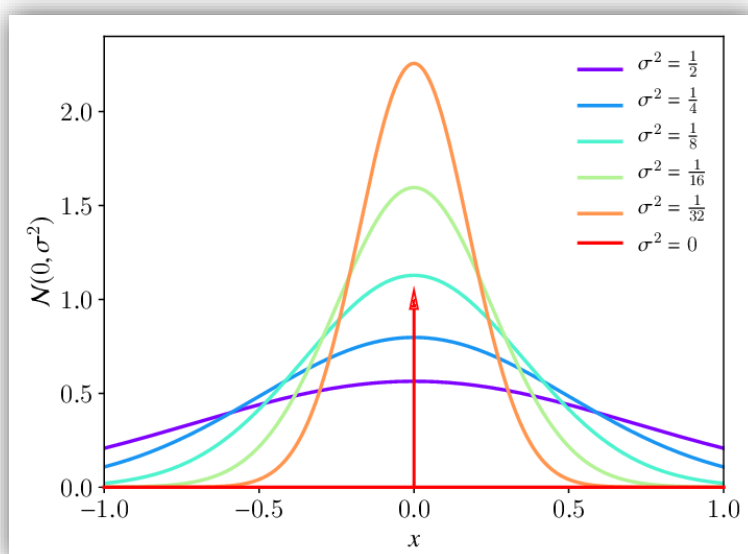
Στο συγκεκριμένο είδος μεθόδου, λαμβάνεται υπόψιν ολόκληρη η περιοχή του προσώπου σαν είσοδος. Στη συνέχεια, σαν γενική πρακτική, επιδιώκεται η εξαγωγή χαρακτηριστικών από το σύνολο της εικόνας, την προβολή τους σε ένα υποχώρο μικρότερης διάστασης ώστε εν τέλη να προχωρήσουμε στην διαδικασία της αναγνώρισης με μέτρηση αποστάσεων μεταξύ των χαρακτηριστικών γνωστών εικόνων και αυτής ως προς αναγνώρισης (Karamizadeh & Abdullah, 2013).

Οι πιο γνωστές είναι η Ανάλυση Γραμμικής Διάκρισης (Linear Discriminate Analysis - LDA), η Ανάλυση κύριων συνιστωσών (Principal Component Analysis - PCA) και η αντίστοιχη για της ανεξάρτητες συνιστώσες (Independent Component Analysis). Για να κατανοήσουμε τα πλεονεκτήματα και τα μειονεκτήματα της συγκεκριμένης ομάδας μεθόδων, είναι σημαντικό να προχωρήσουμε σε μία σύντομη περιγραφή των παραπάνω αλγορίθμων.

Linear Discriminate Analysis (LDA)

Αρχικά, πρέπει να αναφερθεί ότι η Ανάλυση Γραμμικής Διάκρισης, αποτελεί εν γένει, έναν τρόπο μείωσης της διαστατικότητας (dimensionality reduction) που παρατηρείται σε προβλήματα μηχανικής μάθησης, μεγάλων δεδομένων (Big data analysis) και όχι μόνο. Για την σημασία της μικρής διαστατικότητας σε προβλήματα μηχανικής μάθησης έχουμε αναφερθεί παραπάνω και δεν θα προχωρήσουμε στην ανάλυση της σημασίας της τόσο από θέμα απόδοσης του κατηγοροποιητή όσο και κατανάλωσης πόρων. Η συγκεκριμένη τεχνική (Bhattacharyya & Rahul, 2013) προέρχεται από μία άλλη που ονομάζεται Γραμμική Διακρίνουσα (Linear Discriminant) ή αλλιώς Ανάλυση Διακρίνουσας του Fisher (Fisher's Discriminant Analysis) και ο αρχικός της σκοπός ήταν να πετύχει κατηγοριοποίηση σε πρόβλημα δύο μόνο κατηγοριών (binary classification). Στη συνέχεια όμως, κοντά στα μέσα του 20^{ου} αιώνα, πραγματοποιήθηκε επέκτασή της σε πρόβλημα πολλών κατηγοριών. Εδώ, πρέπει να αναφερθεί ότι ένας λόγος για την ανάπτυξή της, εντοπίζεται στο γεγονός ότι παρέχει λύσεις στα προβλήματα που αντιμετώπιζε η λογιστική παλινδρόμηση

(logistic regression) ως μέθοδο κατηγοριοποίησης στην μηχανική μάθηση. Η ίδια, ενώ λειτουργούσε ικανοποιητικά για δυαδικά προβλήματα κατηγοριοποίησης, η επέκτασή της σε πολλαπλές τάξεις, οι περιπτώσεις όπου τα δεδομένα για κάθε κατηγορία ήταν καλώς διαχωρισμένα και τα προβλήματα με μικρά σετ εκπαίδευσης, αποτελούσαν περιπτώσεις όπου η ίδια παρείχε ασταθείς λύσεις. Σαν εισαγωγή, θα μπορούσαμε να πούμε ότι εκμεταλλεύεται αρκετά από τα στατιστικά μεγέθη που μπορούμε να εξάγουμε από ένα σετ δεδομένων όπως η απόκλιση και η μέση τιμή για την κάθε κατηγορία, αφού μετά τον υπολογισμό τους τις χρησιμοποιεί ως εισόδους στις αντίστοιχες εξισώσεις για να πραγματοποιήσει προβλέψεις.



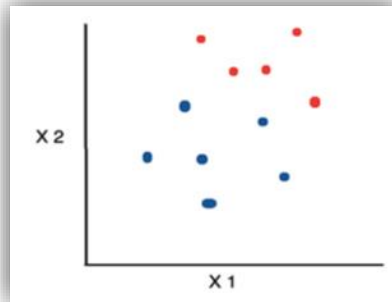
Εικόνα 45: Κανονική Κατανομή

Η LDA προχωράει σε ορισμένες προβλέψεις για το διαθέσιμο σετ δεδομένων.

Αναλυτικά:

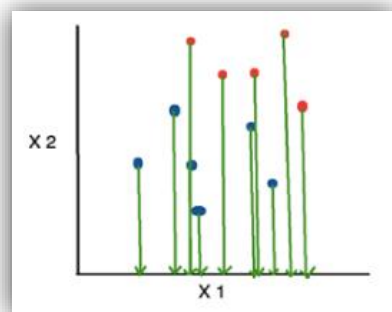
1. Θεωρεί ότι κάθε μεταβλητή ακολουθεί μία γκαουσιανή - κανονική (Gaussian) κατανομή (εικόνα 45)
2. Θεωρεί ότι κάθε χαρακτηριστικό, έχει την ίδια τιμή απόκλισης από την μέση τιμή, κατά μέσο όρο.

Σαν γενική ιδέα του πώς λειτουργεί, έστω το παρακάτω δυσδιάστατο πρόβλημα



Εικόνα 46: Δυσδιάστατο Πρόβλημα

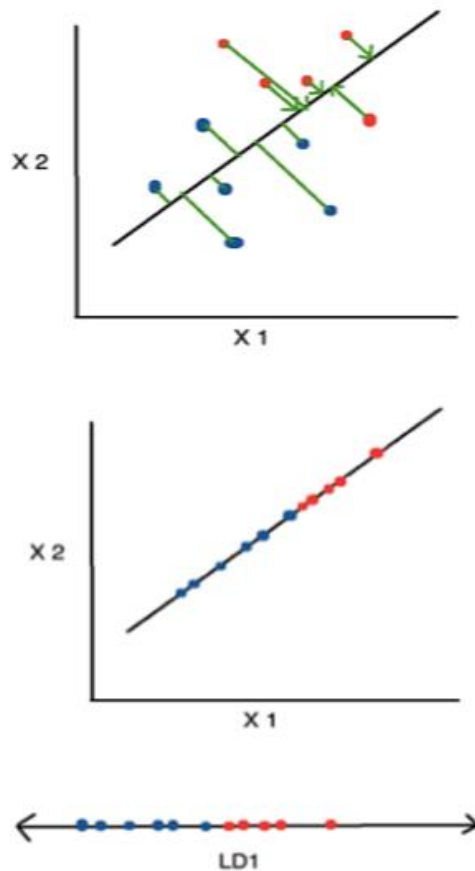
Φαίνεται ότι έχουμε 2 κατηγορίες (μπλε-πορτοκαλί) και έστω ότι θέλουμε να μειώσουμε τις διαστάσεις (δεν ξεχνάμε ότι τα προβλήματα αυτά έχουν ως βασικό στόχο το dimensionality reduction), δηλαδή να προβάλλουμε όλα τα δεδομένα σε έναν άξονα (μιας και έχουμε 2 εδώ). Μία σύντομη αλλά κακή λύση είναι να αφαιρέσεις αυθαίρετα την X_2 διάσταση:



Εικόνα 47: Διατήρηση της μίας διάστασης

Το πρόβλημα είναι, ότι χάνουμε, με αυτόν τον τρόπο σημαντικές πληροφορίες μη όντας ικανοί να διαχωρίσουμε τις δύο κατηγορίες ικανοποιητικά. Σε αντίθεση με αυτήν την τακτική, η LDA επιδιώκει να κατασκευάσει (μέσα από τα δεδομένα) μία καινούργια διάσταση στην οποία, στη συγκεκριμένη περίπτωση να επιτευχθεί και η μείωση των διαστάσεων του προβλήματος και να έχουμε τον μέγιστο δυνατό διαχωρισμό μεταξύ των δύο κατηγοριών.

Αυτό μπορούμε να το δούμε, στο συγκεκριμένο υποθετικό παράδειγμα, παρακάτω:



Εικόνα 48: LDA Λύση

Για την υλοποίηση της LDA, πρέπει να πραγματοποιηθεί ένα σύνολο βημάτων.

Συγκεκριμένα:

1. Υπολογισμός της απόστασης - διαφοράς των μέσων τιμών μεταξύ διαφορετικών κλάσεων. Ονομάζεται αλλιώς και απόκλιση μεταξύ των κλάσεων. Η συγκεκριμένη πράξη εκφράζεται από τον παρακάτω μαθηματικό τύπο:

$$S_b = \sum_{i=1}^g N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

2. Υπολογισμός της απόκλισης στην εκάστοτε κατηγορία, ουσιαστικά υπολογίζοντας την διαφορά μεταξύ της μέσης τιμής μίας κατηγορίας και των δειγμάτων της.

$$S_w = \sum_{i=1}^g (N_i - 1) S_i = \sum_{i=1}^g \sum_{j=1}^{N_i} (x_{i,j} - \bar{x})(x_{i,j} - \bar{x})^T$$

3. Κατασκευή του ζητούμενο χώρου, μειωμένης διάστασης, ο οποίος είναι σε θέση να μεγιστοποιήσει το 1^ο βήμα (δηλαδή την απόσταση των μέσων μεταξύ των κατηγοριών) και να ελαχιστοποιήσει το 2^ο (ώστε όλα τα δείγματα μίας κατηγορίας να βρίσκονται κοντά το ένα με το άλλο). Το συγκεκριμένο κριτήριο, ονομάζεται κριτήριο Fisher και ως P χαρακτηρίζεται η προβολή των σημείων στον νέο χώρο μειωμένων διαστάσεων.

$$P_{lda} = \operatorname{argmax} \frac{P^T S_b P}{P^T S_w P}$$

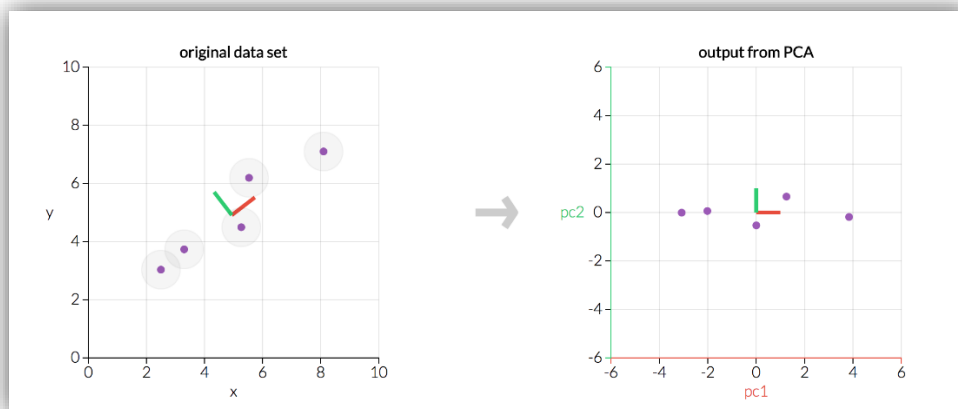
Όσον αφορά, τον τρόπο που πραγματοποιείται εν τέλει η εκτίμηση μίας νέα τιμής για την κατηγορία της, χρησιμοποιείται το θεώρημα του Bayes.

Στον τομέα της αναγνώρισης προσώπου, η παραπάνω στρατηγική, απαιτεί αρχικά ένα σημαντικό ποσό εικόνων με διαφορετικά χαρακτηριστικά προσώπου ως σετ δεδομένων, μία διαδικασία, η ποιότητα της οποίας θα καθορίσει, σε μεγάλο βαθμό και την αξιοπιστία των αποτελεσμάτων. Οι διαφορετικοί αυτοί τύποι προσώπου πρέπει να περιλαμβάνουν ένα σύνολο περιπτώσεων μπροστινής όψης (frontal view) με μικρές τροποποιήσεις ως προς την γωνία λήψης της καθεμίας. Σε συνδυασμό με αυτό, διαφορετικοί τύποι φωτεινότητας, εκφράσεων του προσώπου, background των εικόνων ακόμη και αξεσουάρ, όπως είναι τα γυαλιά μυωπίας, είναι απαραίτητα να υπάρχουν στο σύνολο δεδομένων, τόσο της εκπαίδευσης, όσο και της αξιολόγησης ώστε να παραχθούν αξιόπιστοι κατηγοροποιητές.

Κάθε μία από τις 2D εικόνες, παρουσιάζεται ως ένα διάνυσμα μήκους N στον πολυδιάστατο χώρο. Επίσης έστω ότι έχουμε M δείγματα και C διαφορετικές κατηγορίες. Σύμφωνα με την θεωρία, γνωρίζουμε ότι η LDA θα επιδιώξει να προβάλλει τα διανύσματα αυτά υπό ένα συγκεκριμένο πλαίσιο όπου η διαφορά μεταξύ των μέσων τιμών διαφορετικών κατηγοριών θα είναι μέγιστη (S_b), ενώ η απόκλιση μεταξύ των δειγμάτων μίας τάξης (κλάσης) από την μέση τιμή να είναι η ελάχιστη (S_w).

Principal Component Analysis (PCA)

Μία αντίστοιχα γνωστή μέθοδος για την αντιμετώπιση της ύπαρξης πολυδιαστατικότητας στα προβλήματα μηχανικής μάθησης είναι η Ανάλυση των Κύριων συνιστωσών. Αρχικά, οι μεταβλητές εισόδου υφίσταται μία κανονικοποίηση, μία διαδικασία προ επεξεργασίας αρκετά σημαντική καθώς η ίδια η φύση της PCA είναι αρκετά ευαίσθητη ως προς τις αποκλίσεις των αρχικών (θα εξηγηθεί παρακάτω ο όρος αρχική) παραμέτρων (Paul & Suman, 2012). Αυτό σημαίνει ότι οι τιμές των αριθμητικών αποκλίσεων που είναι μεγαλύτερες (πχ μία τιμή με εύρος τιμών [0,100]) θα κυριαρχήσουν έναντι των μικρότερων (η διαφορά σε μία μεταβλητή που λαμβάνει τιμές στο εύρος [0,1]), γεγονός που ελλοχεύει κινδύνους ως προς την απώλεια πληροφορίας, στην περίπτωση όπου δεν προχωρήσουμε σε standardization ενέργειες.



Εικόνα 49: Τροποποίηση του σετ δεδομένων μέσω της PCA

Ο κλασικός τρόπος για να πραγματοποιηθεί κανονικοποίηση, είναι ο παρακάτω όπου x η τιμή του δείγματος:

$$z = \frac{x - \bar{x}}{std}$$

Έπειτα, προχωράμε στον υπολογισμό του πίνακα συνδυακόμενης (covariance matrix) για όλες τις παραμέτρους εισόδου, με σκοπό να εντοπιστούν εκείνες που αλληλεξαρτώνται σε τέτοιο βαθμό που θα μπορούσαν κάποιες να αφαιρεθούν γιατί περιλαμβάνουν ασήμαντες επιπρόσθετες πληροφορίες. Η εξάρτηση αυτή δεν είναι απαραίτητη να έχει μόνο θετική χροιά, πχ να αυξάνονται και οι δύο μεταβλητές γραμμικά ή μία με τη άλλη αλλά και σε αντιστρόφως ανάλογες περιπτώσεις όπως πχ

το αν έχει ήλιο έξω ή όχι (η μία περίπτωση αποκλείει την άλλη οπότε θα μπορούσαμε να συμπτύξουμε τις δύο παραμέτρους σε μία). Για παράδειγμα, έστω ότι έχουμε ένα σετ δεδομένων με 3 χαρακτηριστικά, δηλαδή τρισδιάστατο. Τότε ο πίνακας συνδυακόμενης θα ήταν ο εξής:

$$\begin{vmatrix} Cov(x, x) & Cov(x, y) & Cov(x, z) \\ Cov(y, x) & Cov(y, y) & Cov(y, z) \\ Cov(z, x) & Cov(z, y) & Cov(z, z) \end{vmatrix}$$

Για να υπολογίσουμε τον εκάστοτε όρο, πρέπει πρώτα να λάβουμε υπόψιν ότι ο τύπος της τυπικής απόκλισης από την μέση τιμή για μία είσοδο-χαρακτηριστικό-feature x υπολογίζεται με τον παρακάτω τύπο, με αριθμό δειγμάτων n :

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

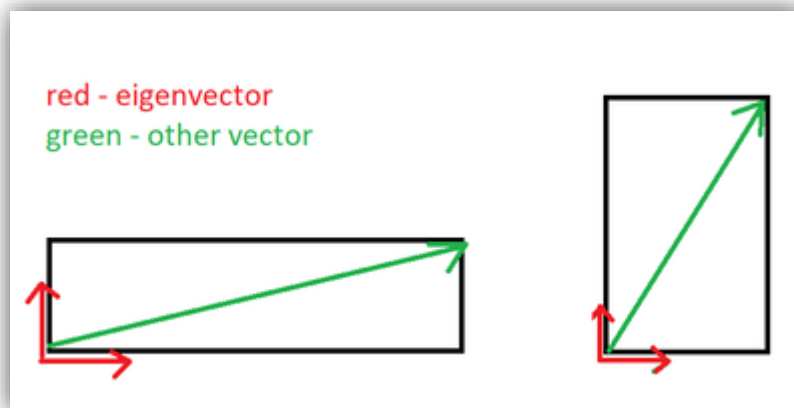
Στην περίπτωση όπου αναζητάμε την συνδιακόμευση μεταξύ μίας μεταβλητής και του εαυτού της (δηλαδή τις τιμές στις θέσεις (0,0), (1,1) και (2,2) του παραπάνω πίνακα (στις διαγώνιες θέσεις), τότε, στην ουσία βρίσκουμε την τυπική απόκλιση που αναφέραμε πριν. Στην περίπτωση όμως, που θέλουμε να βρούμε την συνδιακόμευση μεταξύ δύο διαφορετικών μεταβλητών, τότε ο τύπος που θα χρησιμοποιήσουμε είναι ο παρακάτω:

$$\sigma(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Η σημασία της έννοιας της συνδιακόμενης μπορεί να εντοπιστεί στο πρόσημο, όπου σε περίπτωση που έχουμε θετικό, τότε δηλώνεται μία θετική συσχέτιση μεταξύ των δύο μεταβλητών, ενώ σε περίπτωση αρνητικού πρόσημου, μία αντιστρόφως ανάλογη εξάρτηση.

Όσον αφορά την αναγνώριση προσώπου, οι PCA προσεγγίζεις, στην πλειονότητά τους, αξιοποιούν μία μέθοδο ιδιοδιανυσμάτων, που στην περίπτωση της αναγνώρισης προσώπου, ονομάζεται Eigenface (eigenvectors – face) ή αλλιώς χαρακτηριστικό πρόσωπο ή ‘ιδιοπρόσωπο’. Για να είμαστε σε θέση να κατανοήσουμε σε βάθος τον πυρήνα της μεθόδου, πρέπει να αναλυθεί η έννοια του χαρακτηριστικού διανύσματος και της αντίστοιχης τιμής – ιδιοτιμής.

Το ιδιοδιάνυσμα (eigenvector) αποτελεί ένα διάνυσμα του οποίου η κατεύθυνση (διεύθυνση – φορά) παραμένει ανεπηρέαστη, στην περίπτωση όπου το εφαρμοστεί κάποιος γραμμικός μετασχηματισμός.



Εικόνα 50: Ιδιοδιανύσματα - Παράδειγμα γραμμικού μετασχηματισμού

Έστω, το παραπάνω παράδειγμα γραμμικού μετασχηματισμού, όπου ο αντίστοιχος πίνακας είναι ο εξής:

$$A = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}$$

Αυτό που αναφέρει ο πίνακας, είναι ότι μέσω του συγκεκριμένου μετασχηματισμού μειώνουμε στο μισό το μήκος στον άξονα x και διπλασιάζουμε το μήκος στον άξονα y. Έστω ακόμη, ένα διάνυσμα:

$$\vec{u} = (x, y) \text{ και } \vec{u'} = A\vec{u}$$

Εύκολα μπορούμε να παρατηρήσουμε, εδώ ότι ενώ θα επηρεαστούν, από θέμα αριθμητικής τιμής, οι κατευθύνσεις των ιδιοδιανυσμάτων (κόκκινα βέλη) δεν αλλάζουν, σε σχέση με τα άλλα (το πράσινο βέλος). Για έναν τετραγωνικό πίνακα A, τα ιδιοδιανύσματα που τον ορίζουν, μοναδικά (για αυτό και ονομάζονται και χαρακτηριστικά), είναι εκείνα που ικανοποιούν την παρακάτω σχέση:

$$A\vec{u} = \lambda\vec{u}$$

όπου λ ονομάζεται ιδιοτιμή.

Συνεχίζοντας λίγο την εξίσωση:

$$\vec{u}(A - \lambda I) = 0 \text{ και } \text{Det}(A - \lambda I) = 0$$

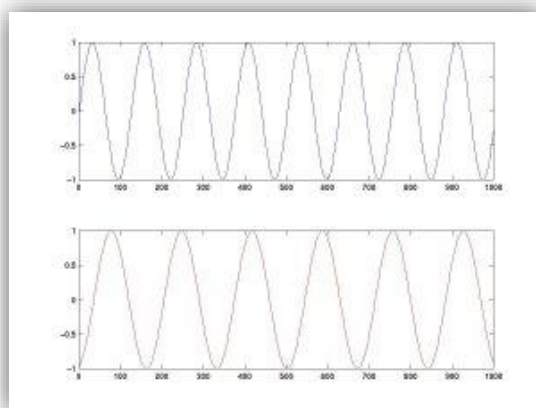
Στην περίπτωση όπου έχουμε μία εικόνα 32x32 pixels και θέλουμε να αντιμετωπίσουμε τις τιμές της φωτεινότητας κάθε pixel ως χαρακτηριστικό, κατανοούμε ότι μπορούμε να παράγουμε ένα διάνυσμα 1024 χαρακτηριστικών – features. Η ανάγκη για μείωση της διαστατικότητας, φαίνεται στους υπολογιστικούς πόρους αν πχ προσπαθήσουμε να εκτελέσουμε την μέθοδο της ευκλείδειας απόστασης για χιλιάδες εικόνες για κάθε pixel. Στην PCA αυτό που κάνουμε είναι να πετυχαίνουμε την μείωση του χώρου των χαρακτηριστικών, υπολογίζοντας τα ιδιοδιανύσματα του πίνακα συνδιακύμανσης αυτών των διαστάσεων και προβάλλοντας, στη συνέχεια κάθε διάνυσμα του χαρακτηριστικού στο μεγαλύτερο ιδιοδιάνυσμα.

Για παράδειγμα, λαμβάνοντας ως δεδομένο ότι τα ιδιοδιανύσματα ενός 2D προβλήματος είναι δισδιάστατα (x,y) στην περίπτωση του 1024, θα έχουν 1024 διαστάσεις. Κάθε feature, σύμφωνα με αυτά που ειπώθηκαν παραπάνω, μπορεί να προβληθεί πάνω στα N μεγαλύτερα ιδιοδιανύσματα, των οποίων ο γραμμικός συνδυασμός μπορεί να το αναπαραστήσει. Τοποθετώντας τα κατάλληλα βάρη σε έναν συγκεκριμένο αριθμό των πρώτων διανυσμάτων (έστω μία τιμή 70) είμαστε σε θέση να διατηρήσουμε την μεγαλύτερη ποσότητα της πληροφορίας της εικόνας, μειώνοντας σημαντικά την διαστατικότητα. Για το πρόβλημα του υπολογισμού των ιδιοπροσώπων, δεν υπάρχει συγκεκριμένος αριθμός, αλλά περισσότερο εμπειρικές επιλογές και trial and error διαδικασίες.

Στη συνέχεια θα προχωρήσουμε στην ανάλυση της τελευταίας μεθόδου για την συγκεκριμένη κατηγορία model – based προσεγγίσεων, την Ανάλυση Ανεξάρτητων Συνιστωσών (Independent Component Analysis).

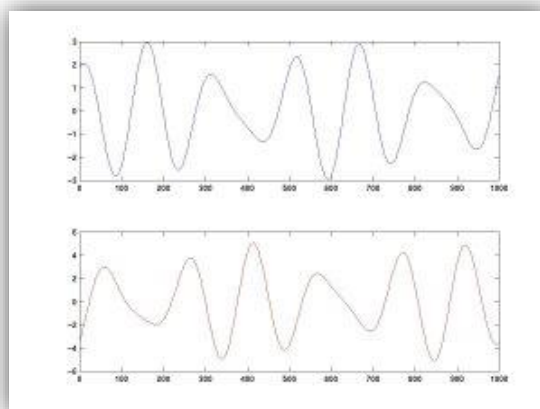
Individual Component Analysis (ICA)

Η ICA αποτελεί μία μέθοδο που επιδιώκει τον διαχωρισμό γραμμικά αναμειγμένων πηγών (Bartlett, Movellan, & Sejnowski, 2010). Η λειτουργία της την έχει κάνει ιδιαίτερα διάσημη στον τομέα της επεξεργασίας σήματος, καθώς όπως φαίνεται παρακάτω, έστω ότι έχουμε δύο ανεξάρτητες πηγές:



Εικόνα 51: Ανεξάρτητες Πηγές Σήματος

Και, έπειτα τις συνθέτουμε γραμμικά με δύο διαφορετικούς γραμμικούς τρόπους, δηλαδή $xA+yB$, έτσι ώστε να πάρουμε το παρακάτω αποτέλεσμα:



Εικόνα 52: Γραμμικός Μετασχηματισμός των δύο πηγών

Με την ICA είμαστε σε θέση να τις ανακατασκευάσουμε από τους γραμμικούς μετασχηματισμούς της. Για να πραγματοποιηθεί αυτό, λαμβάνουμε ως δεδομένο ότι τα δύο σήματα δεν ακολουθούν, αρχικά μία γκαουσιανή κατανομή. Έπειτα, είναι σημαντικό τα δύο σήματα να είναι στατιστικά ανεξάρτητα. Για αυτό μπορεί να λάβει

χώρα ένας γραμμικός μετασχηματισμός του διανύσματος των τυχαίων μεταβλητών με ένα γνωστό πίνακα συνδιακύμανσης σε ένα νέο σετ μεταβλητών όπου ο ίδιος ο πίνακας θα είναι μοναδιαίος. Αυτό, σημαίνει να είμαστε σε θέση να αφαιρέσουμε οποιαδήποτε συσχέτιση μεταξύ των μεταβλητών, ώστε πλέον το καινούργιο σύνολο να είναι ασυσχέτιστο με απόκλιση 1 από την μέση τιμή (Whitening – Sphere μετασχηματισμός). Ως αποτέλεσμα για τα δύο αυτά σήματα x και y μπορούμε να γράψουμε:

$$P(x \cap y) = P(x)P(y)$$

Στη συνέχεια, αν πολλαπλασιάσουμε τα δύο αυτά σήματα με ένα κατάλληλο πίνακα ώστε να ελαχιστοποιήσουμε την γκαουσιανή κατανομή των δεδομένων τους, είμαστε σε θέση μέσω του θεωρήματος κεντρικού ορίου, το οποίο λέει ότι αν προσθέσουμε δύο ανεξάρτητες πηγές, το άθροισμά τους τείνει να φέρει περισσότερο σε κανονική κατανομή από την κάθε πηγή ξεχωριστά, να βρούμε τις δύο αρχικά ανεξάρτητες πηγές.

Στις εφαρμογές της αναγνώρισης προσώπου, ακολουθείται η ίδια λογική της ανάμειξης των σημάτων ώστε στη συνέχεια να προχωρήσουμε στον διαχωρισμό τους. Συγκεκριμένα, οι εικόνες που τοποθετούνται σαν είσοδο θεωρούνται ότι είναι μία γραμμική μίξη από στατιστικά ανεξάρτητες εικόνες που λειτουργούν ως ‘βάση’ σε συνδυασμό με έναν άγνωστο πίνακα A .

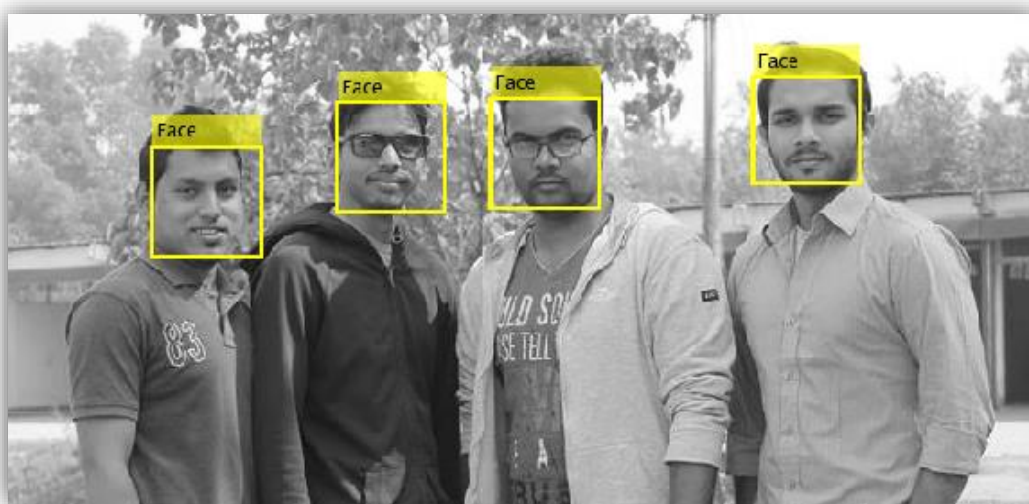
Όπως αναφέρθηκε αυτές οι 3 μέθοδοι, χρησιμοποιούνται σε μία πληθώρα εφαρμογών και η αναγνώριση προσώπου δεν αποτελεί εξαίρεση. Παρόλα αυτά όμως, παρουσιάζουν μειονεκτήματα, καθώς όπως φαίνεται, λαμβάνεται όλη η εικόνα ως είσοδος και πραγματοποιούνται αριθμητικές πράξεις μεταξύ pixels γεγονός που αυξάνει σημαντικά τους υπολογιστικούς πόρους και την απόδοση συστημάτων που βασίζονται σε μία model based τεχνική, καθιστώντας μη βέλτιστη για εφαρμογές πραγματικού χρόνου.

Feature Based Κατηγοριοποίηση

Στο συγκεκριμένο είδος των συστημάτων, οι ερευνητές εκμεταλλεύονται το γεγονός ότι το ανθρώπινο πρόσωπο φέρει ορισμένα χαρακτηριστικά (features) τα οποία είναι καθολικά, όπως τα φρύδια, η μύτη και τα μάτια. Με αυτή την υπόθεση, είναι δυνατόν να αναπτυχθούν αλγόριθμοι οι οποίοι θα είναι πολύ ταχύτεροι από την προηγούμενη κατηγορία ενώ ταυτόχρονα θα είναι σε θέση να παρέχουν αξιόπιστες προβλέψεις. Ο πιο γνωστός αλγόριθμος για εντοπισμό προσώπου (πρόσοψη) που χρησιμοποιείται ακόμη και σήμερα από τα κινητά και τις webcams είναι ο Viola – Jones (Viola & Jones, Michael, 2001).

Viola – Jones Αλγόριθμος

Το πρώτο βήμα είναι η μετατροπή της εικόνας προς αναγνώριση σε greyscale, καθώς με αυτόν τον τρόπο μειώνεται σημαντικά το υπολογιστικό κόστος. Στη συνέχεια, μετά την αναγνώριση του προσώπου, θα πραγματοποιηθεί αντιστοίχιση στην εικόνα με τα χρώματα. Ο συγκεκριμένος αλγόριθμος θα ‘καταλήξει’ στον σχεδιασμό ενός κουτιού το οποίο θα σηματοδοτεί το πρόσωπο που έχει αναγνωριστεί.

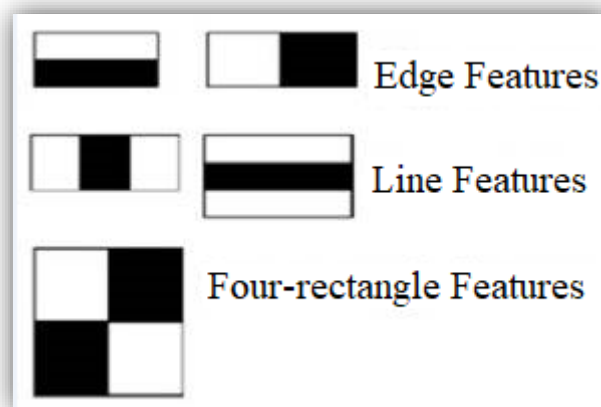


Εικόνα 53: Σχεδιασμός του box – Viola Jones

Μέχρι, βέβαια να υπάρξει αυτό το αποτέλεσμα, ο συγκεκριμένος αλγόριθμος ψάχνει με αυτό το τετράγωνο – κουτί σε όλη την εικόνα, μετακινώντας το κάθε φορά από αριστερά προς τα δεξιά αυτής (το μέγεθος του κουτιού μπορεί να διαφέρει), αναζητώντας ένα πρόσωπο ή αλλιώς τα απαιτούμενα haar – χαρακτηριστικά. Για να

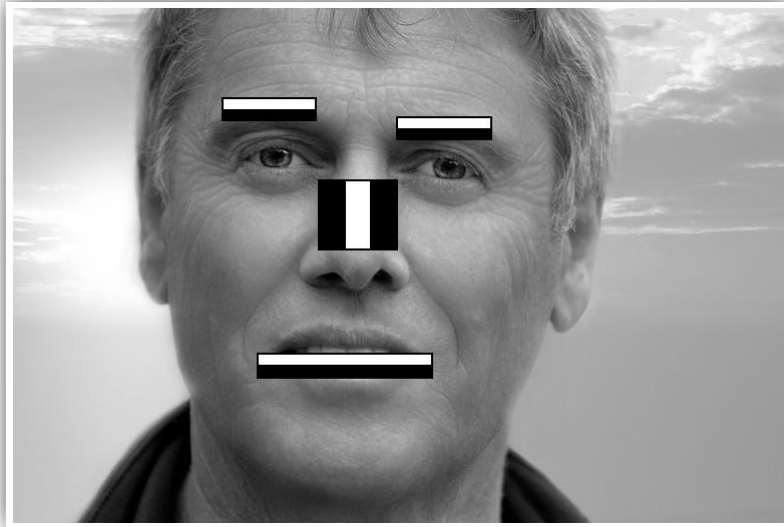
κατανοήσουμε καλύτερα πώς δουλεύουν, είναι σημαντικό να κάνουμε μία εισαγωγή και στο τι είναι τα haar χαρακτηριστικά.

Τα Haar χαρακτηριστικά αναπτύχθηκαν τον 19^ο αιώνα από τον Ούγγρο μαθηματικό Alfred Haar, από τον οποίο πήραν και το όνομά τους. Τα συγκεκριμένα αποτελούνται από ένα κουτί που έχει μία φωτεινή και μία τελείως σκοτεινή πλευρά, με σκοπό να εντοπιστεί κάποιο συγκεκριμένο χαρακτηριστικό ενός προσώπου, που σε μία greyscale εικόνα πράγματι θα παρουσιάζει μία όμοια συμπεριφορά όπως πχ το φρύδι ή το μάτι.



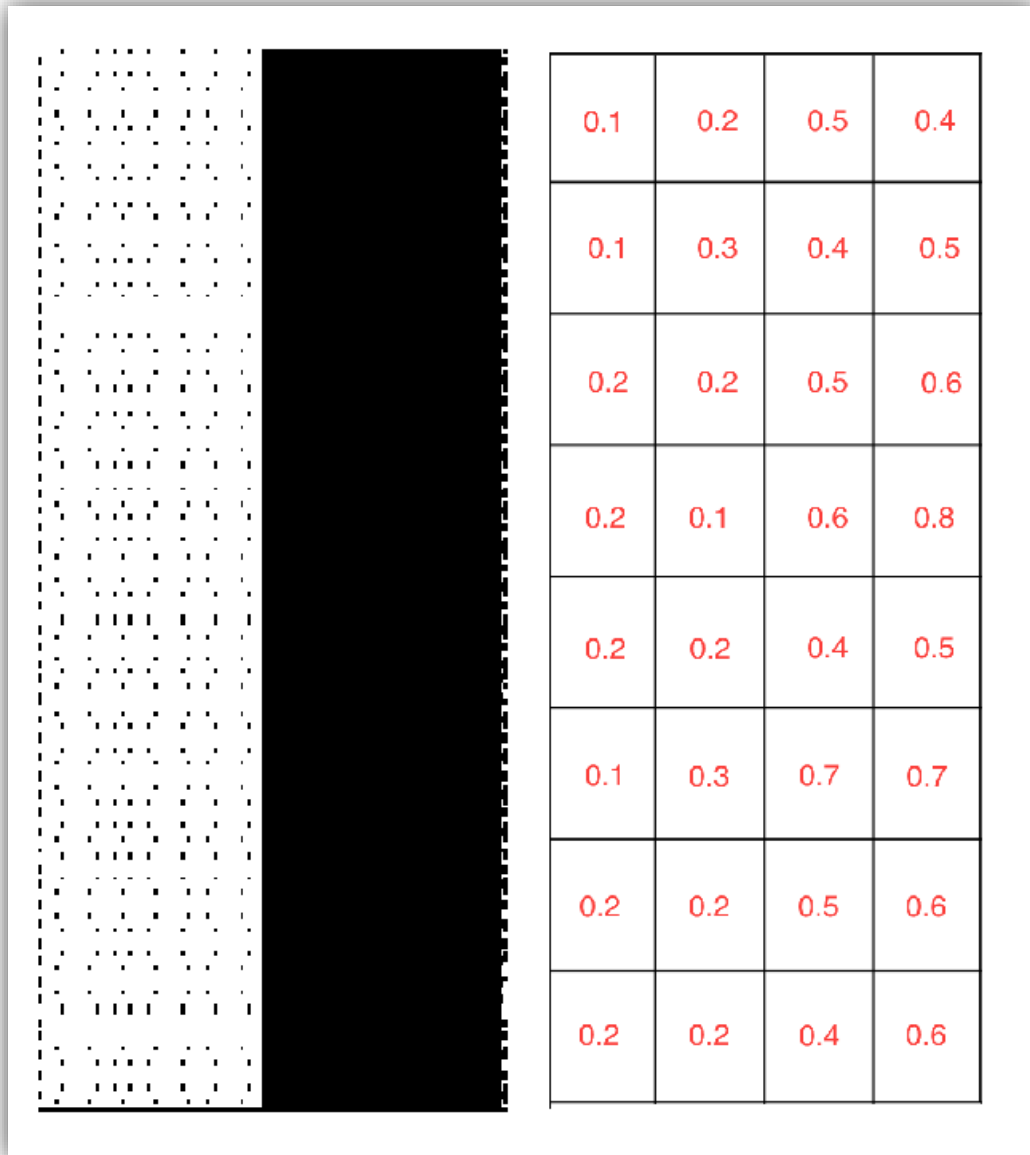
Εικόνα 54: Haar Χαρακτηριστικά

Υπάρχουν 3 είδη Haar χαρακτηριστικών όπως φαίνεται και από την εικόνα 54 και παρακάτω φαίνεται πώς σε μία ασπρόμαυρη εικόνα είναι δυνατόν να εντοπιστούν:



Εικόνα 55: Εντοπισμός Haar χαρακτηριστικών σε μία εικόνα

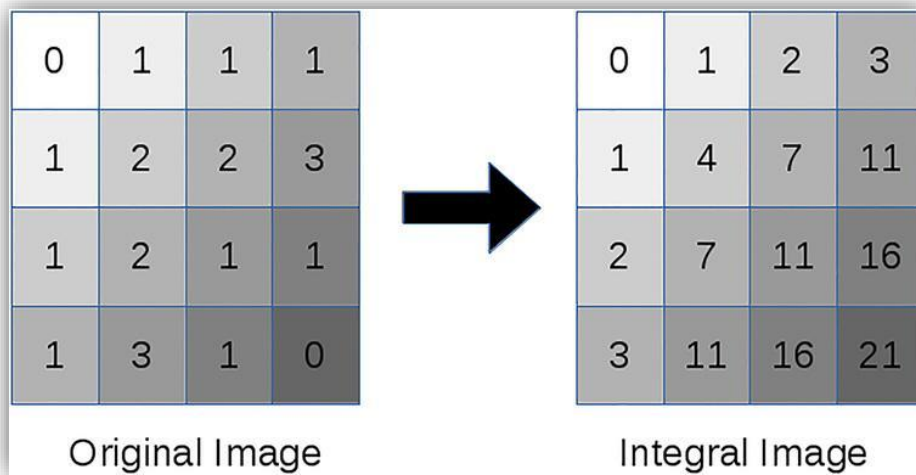
Από την εικόνα 55 φαίνεται λοιπόν ότι τα Edge χαρακτηριστικά μπορούν να βοηθήσουν στον εντοπισμό των φρυδιών και του στόματος ενώ τα Line στην μύτη κτλ. Ως συνέχεια, ο συγκεκριμένος αλγόριθμος αντιμετωπίζει το συγκεκριμένο κουτί-πλαίσιο ως ένα σύνολο τιμών pixel και του δίνει τιμές σχετικές με την φωτεινότητα. Συγκεκριμένα, το εύρος των τιμών κυμαίνεται μεταξύ 0 και 1, όπου όσο πιο σκοτεινό είναι τόσο μεγαλύτερη τιμή έχει.



Εικόνα 56: Σύγκριση των τιμών της φωτεινότητας

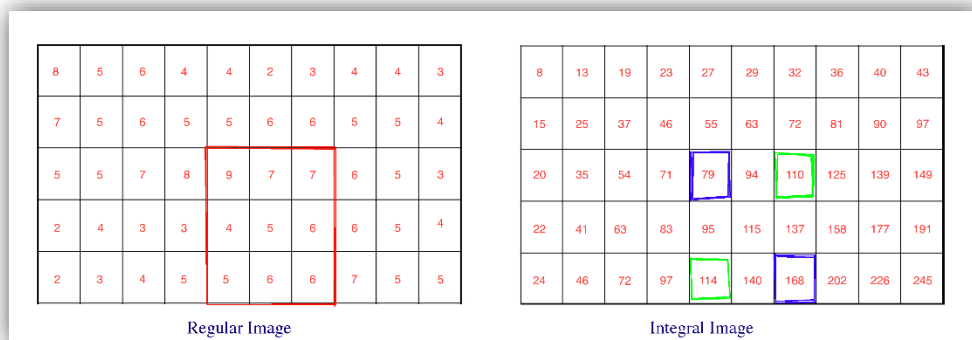
Στη συνέχεια, με βάση αυτές τις μετρήσεις επιδιώκουμε να υπολογίσουμε την διαφορά της φωτεινότητας μεταξύ των δύο περιοχών και αν αυτή η διαφορά, βρίσκεται πάνω από ένα συγκεκριμένο όριο (threshold) τότε το συγκεκριμένο Haar χαρακτηριστικό θεωρείται ότι έχει αναγνωρισθεί. Η τιμή αυτή ονομάζεται τιμή του χαρακτηριστικού (feature) αλλά, είναι κατανοητό ότι αν πραγματοποιηθεί για το σύνολο της εικόνας, τότε θα δημιουργήσει προβλήματα σχετικά με την απόδοση, ειδικά σε μεγαλύτερες διαστάσεις όπου το μέγεθος των pixels θα είναι πολύ μεγαλύτερο. Για τον λόγο αυτό οι ερευνητές που ανέπτυξαν τον παραπάνω αλγόριθμο

το 2001, εισήγαγαν και την έννοια της εικόνας ολοκλήρωσης (integral image). Η ίδια μας επιτρέπει γρήγορα να κατανοήσουμε αν ένα χαρακτηριστικό ενός σετ χαρακτηριστικών πληροί τα απαραίτητα κριτήρια. Κατά την ίδια, έστω μία εικόνα με τις παρακάτω τιμές (έστω ότι τώρα το εύρος δεν είναι μεταξύ 0 και 1), και εμείς θέλουμε να υπολογίσουμε την τιμή της integral εικόνας. Για να το κάνουμε αυτό, προσθέτουμε, σε κάθε pixel τις τιμές όλων αυτών που βρίσκονται στα αριστερά του, όπως φαίνεται παρακάτω:



Εικόνα 57: Κατασκευή της Integral Εικόνας

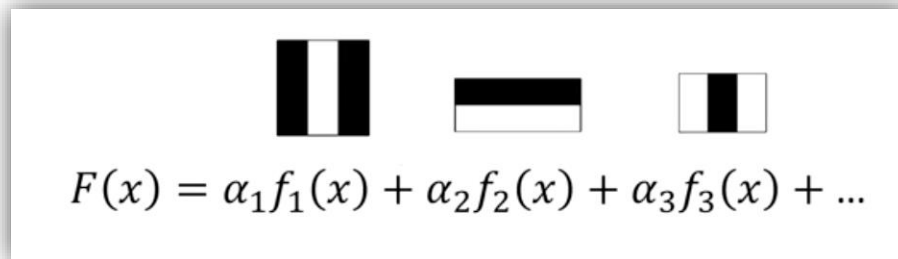
Για το τετράγωνο – box που μας ενδιαφέρει, αρκεί να αφαιρέσουμε από την τιμή της κάτω δεξιά γωνίας, την κάτω αριστερά και να προσθέσουμε το αποτέλεσμα, στο αποτέλεσμα της αφαίρεσης της τιμής της πάνω δεξιάς γωνίας από την πάνω αριστερά.



Εικόνα 58: Υπολογισμός της τιμής του feature του ROI

Ένα σημαντικό ζήτημα, που προηγείται της αναγνώρισης προσώπου είναι η εκπαίδευση των αντίστοιχων κατηγοροποιητών. Αυτό συμβαίνει ώστε να τεθούν τα σωστά όρια – thresholds στις τιμές των pixel για να είναι σε θέση ο αντίστοιχος αλγόριθμος να λειτουργήσει σωστά. Πριν προχωρήσουμε στην υλοποίηση του συγκεκριμένου αλγορίθμου στην γλώσσα python, είναι σημαντικό να μιλήσουμε και για την AdaBoost/Cascading διαδικασίες που βοηθούν τόσο στην απόδοση από θέμα υπολογιστικών πόρων όσο και στην ακρίβεια του μοντέλου.

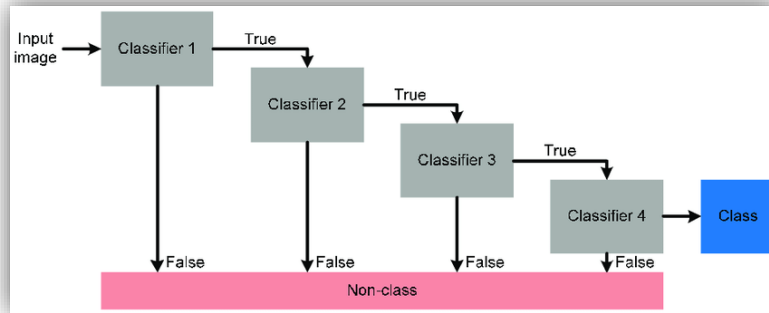
Ο αλγόριθμος Adaptive Boosting ή αλλιώς AdaBoost, επιδιώκει μέσα από τον γραμμικό συνδυασμό των Haar features που υπολογίζονται για κάθε ROI, με την χρήση των κατάλληλων βαρών (weights) να συνδυάσει όλους αυτούς τους αδύναμους κατηγοροποιητές (weak classifiers) σε έναν πιο ισχυρό (strong classifier)



$$F(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \dots$$

Εικόνα 59: Συντελεστές για τον κάθε weak classifier και η δημιουργία του ισχυρού

Από την άλλη, η cascading διαδικασία έχει ήδη αναφερθεί παραπάνω και αφορά το γεγονός ότι κατά τον αλγόριθμο, σε περίπτωση που στην αλληλουχία της αναγνώρισης των Haar χαρακτηριστικών, εντοπιστεί κάποιο που δεν πληροί τις προϋποθέσεις για το συγκεκριμένο τετράγωνο, τότε ολόκληρο το box απορρίπτεται και ο προχωράμε στο επόμενο.



Εικόνα 60: Cascading διαδικασία

Στην ρυθον γράψαμε τον παρακάτω κώδικα:

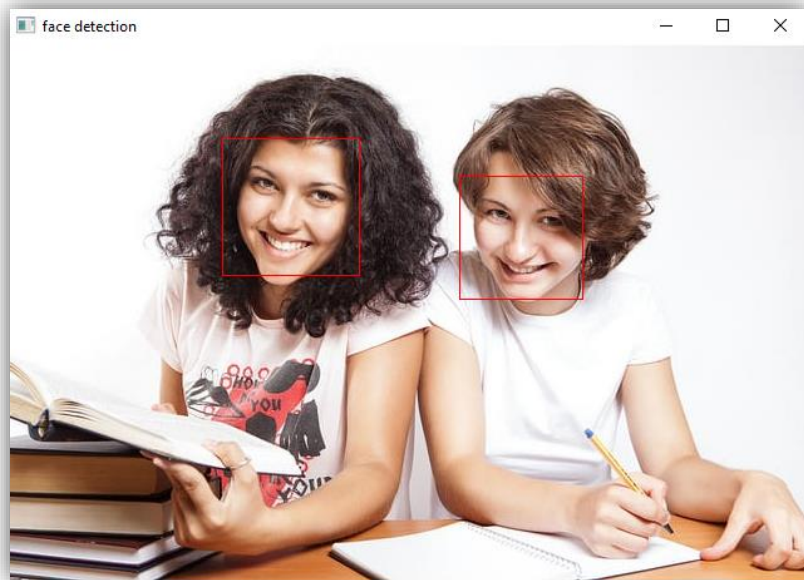
```

import cv2
from cv2 import imread
from cv2 import imshow
from cv2 import waitKey
from cv2 import destroyAllWindows
from cv2 import CascadeClassifier
from cv2 import rectangle
pixels = imread('test2.jpg')
classifier = CascadeClassifier('haarcascade_frontalface_default.xml')
bboxes = classifier.detectMultiScale(pixels, 1.05, 8)
for box in bboxes:
    x, y, width, height = box
    x2, y2 = x + width, y + height
    rectangle(pixels, (x, y), (x2, y2), (0,0,255), 1)
imshow('face detection', pixels)
waitKey(0)
destroyAllWindows()
  
```

Με τον οποίο αξιοποιώντας τις έτοιμες βιβλιοθήκες της OpenCV και τον αλγόριθμο του Viola – Jones, επιδιώκουμε να αναγνωρίσουμε τα πρόσωπα που υπάρχουν στις παρακάτω φωτογραφίες:

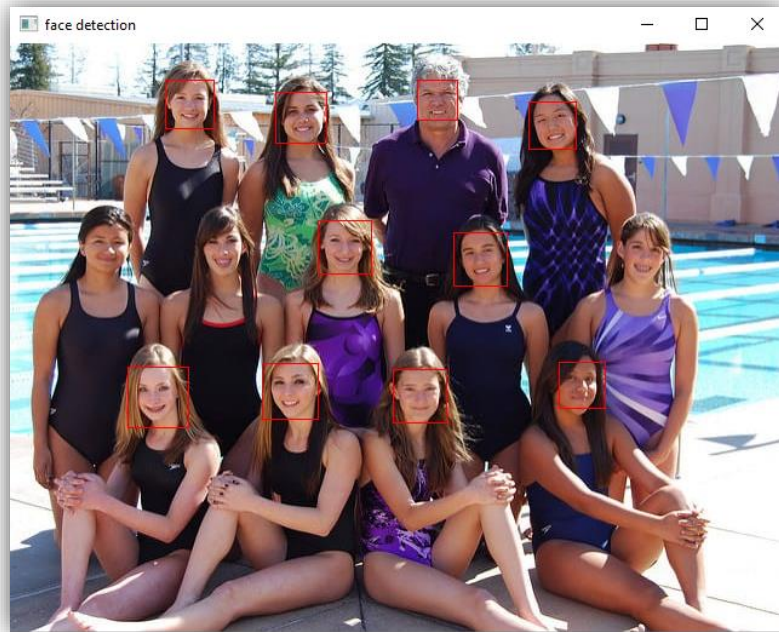


Τρέχοντας τον αλγόριθμο (η εκπαίδευση έχει ήδη πραγματοποιηθεί και το παίρνουμε έτοιμο από την αντίστοιχη βιβλιοθήκη), παρατηρούμε ότι είναι σε θέση να διακρίνει τα πρόσωπα στην 3^η φωτογραφία:



Εικόνα 61: Αποτελέσματα Εικόνας 1

Ενώ στις υπόλοιπες δύο λόγω του γεγονότος ότι περιέχονται περισσότερα πρόσωπα, παρουσιάζει μη βέλτιστα αποτελέσματα



Εικόνα 62: Αποτελέσματα Εικόνας 2

Ενώ, όπως φαίνεται και στα αποτελέσματα της Εικόνας 3, οδηγείται και σε λανθασμένη αναγνώριση προσώπου (κάτω αριστερά εντοπίζεται ο ώμος σαν πρόσωπο):



Εικόνα 63: Αποτελέσματα Εικόνας 3

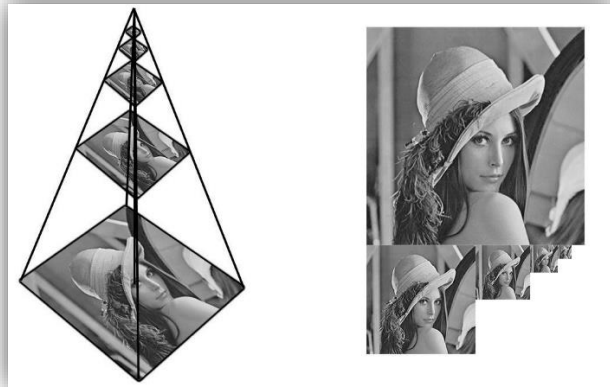
Τα συγκεκριμένα αποτελέσματα, δεν είναι τελικά καθώς είναι δυνατόν να πειραματιστούμε με μία πληθώρα παραμέτρων όπως το μέγεθος του παραθύρου κτλ. Ως αποτέλεσμα, θα μπορούσαμε να αναφέρουμε ότι οι feature based τεχνικές έχουν ως κύριο μέλημα την παραγωγή αποδοτικών αποτελεσμάτων παρόλο το γεγονός ότι οι αντίστοιχοι κατηγοροποιητές, κατά πάσα πιθανότητα, δεν θα φέρουν τις βέλτιστες αποδόσεις. Για τον λόγο αυτό, εισάγονται, λόγω και της ανάπτυξής τους την τελευταία δεκαετία τα νευρωνικά δίκτυα και ειδικότερα τα βαθιά νευρωνικά δίκτυα.

Νευρωνικά Δίκτυα

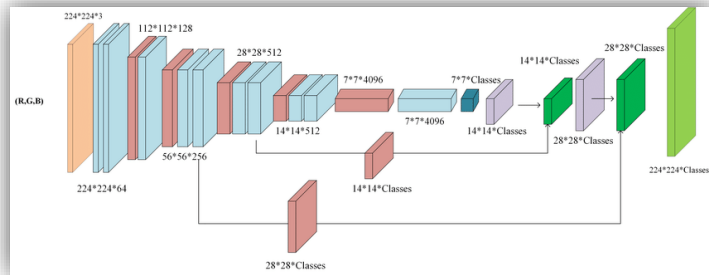
Η θεωρία πίσω από τις συγκεκριμένες αρχιτεκτονικές και σημαντικές έννοιες αυτής έχουν αναλυθεί εκτενώς σε όλη την έκταση της συγκεκριμένης εργασίας οπότε δεν θα προχωρήσουμε σε μεγαλύτερη διερεύνησή τους. Θα περιοριστούμε στην ανάλυση ενός αρκετά διάσημου συστήματος που προτάθηκε το 2016, με τίτλο “Multi-Task Cascaded Convolutional neural Network” ή αλλιώς MTCNN καθώς κατάφερε να πετύχει υψηλές αποδόσεις σε ένα σύνολο εγκαθιδρυμένων σετ δεδομένων καθώς και επειδή είναι σε θέση να αναγνωρίσει συγκεκριμένα χαρακτηριστικά του προσώπου όπως τα μάτια και το στόμα, μία διαδικασία που ονομάζεται landmark detection (Zhang, Zhang, Li, & Qiao, 2016).

Στο MTCNN, υπάρχουν τρία βασικά στάδια:

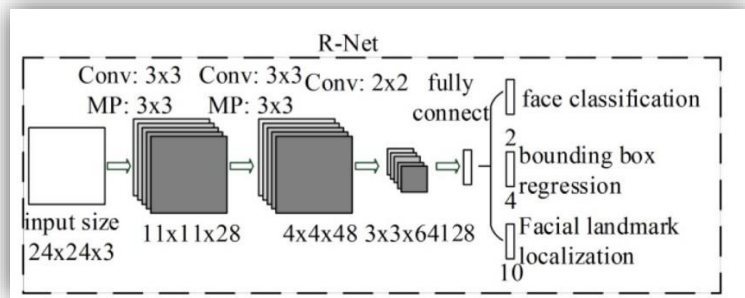
1. Πραγματοποιείται ένα rescaling της εικόνας σε ένα εύρος διαφορετικών μεγεθών (πυραμίδα) το οποίο εύρος θα αποτελέσει την είσοδο ολόκληρης αυτής της αρχιτεκτονικής.
2. Σαν πρώτο δίκτυο (P- Net), υπάρχει ένα πλήρως συνελκτικό δίκτυο (fully convolutional network – FCN) με σκοπό να αποκτήσει ορισμένα υποψήφια παράθυρα και τα αντίστοιχα boxes.
3. Όλες οι υποψήφιες εξοδοί του P- Net αξιοποιούνται σαν είσοδο στο Refine δίκτυο, μία CNN νευρωνική αρχιτεκτονική η οποία επιδιώκει να μειώσει ακόμη περισσότερο τα υποψήφια boxes ενώ ταυτόχρονα να επιτύχει την σύμπτυξη των υποψηφίων παραθύρων που επικαλύπτονται (overlapping). Σε αυτό το στάδιο πραγματοποιείται η απόφαση του αν τελικά αναφερόμαστε σε πρόσωπο ή όχι σε αυτή την εικόνα καθώς και παρέχεται ένα διάνυσμα 10 στοιχείων για το landmark detection που αναφέρθηκε πριν.
4. Το δίκτυο εξόδου (Output Network O-Net) αποτελεί μία αρχιτεκτονική ίδια με αυτή του προηγούμενου σταδίου και είναι υπεύθυνη για την περιγραφή, με μεγαλύτερη λεπτομέρεια των χαρακτηριστικών των προσώπων, καθώς και για την παροχή των θέσεων στην εικόνα για τα μάτια, την μύτη και το στόμα.



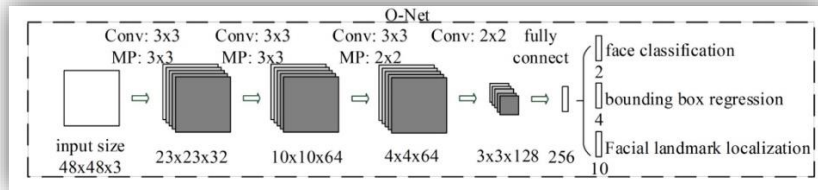
Εικόνα 64: Rescaling της εικόνας



Εικόνα 65: P - Net



Εικόνα 66: R - Net



Εικόνα 67: O – Net

```

# face detection with mtcnn on a photograph
from matplotlib import pyplot
from matplotlib.patches import Rectangle
from matplotlib.patches import Circle
from mtcnn.mtcnn import MTCNN

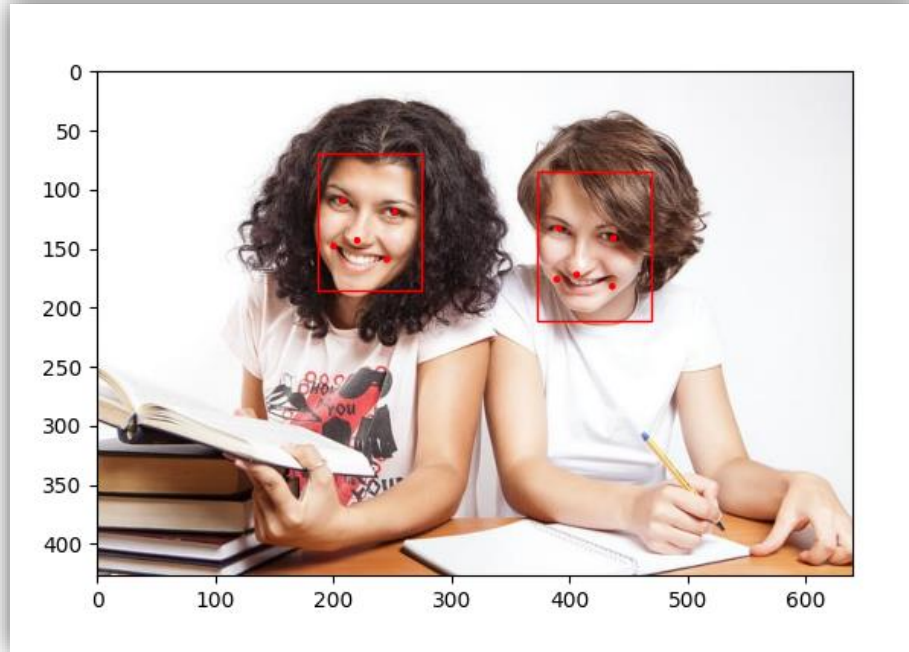
# draw an image with detected objects
def draw_image_with_boxes(filename, result_list):
    # load the image
    data = pyplot.imread(filename)
    # plot the image
    pyplot.imshow(data)
    # get the context for drawing boxes
    ax = pyplot.gca()
    # plot each box
    for result in result_list:
        # get coordinates
        x, y, width, height = result['box']
        # create the shape
        rect = Rectangle((x, y), width, height, fill=False,
color='red')
        # draw the box
        ax.add_patch(rect)
        # draw the dots
        for key, value in result['keypoints'].items():
            # create and draw dot
            dot = Circle(value, radius=2, color='red')
            ax.add_patch(dot)
    # show the plot
    pyplot.show()

filename = 'test3.jpg'
# load image from file
pixels = pyplot.imread(filename)
# create the detector, using default weights
detector = MTCNN()
# detect faces in the image
faces = detector.detect_faces(pixels)
# display faces on the original image
draw_image_with_boxes(filename, faces)

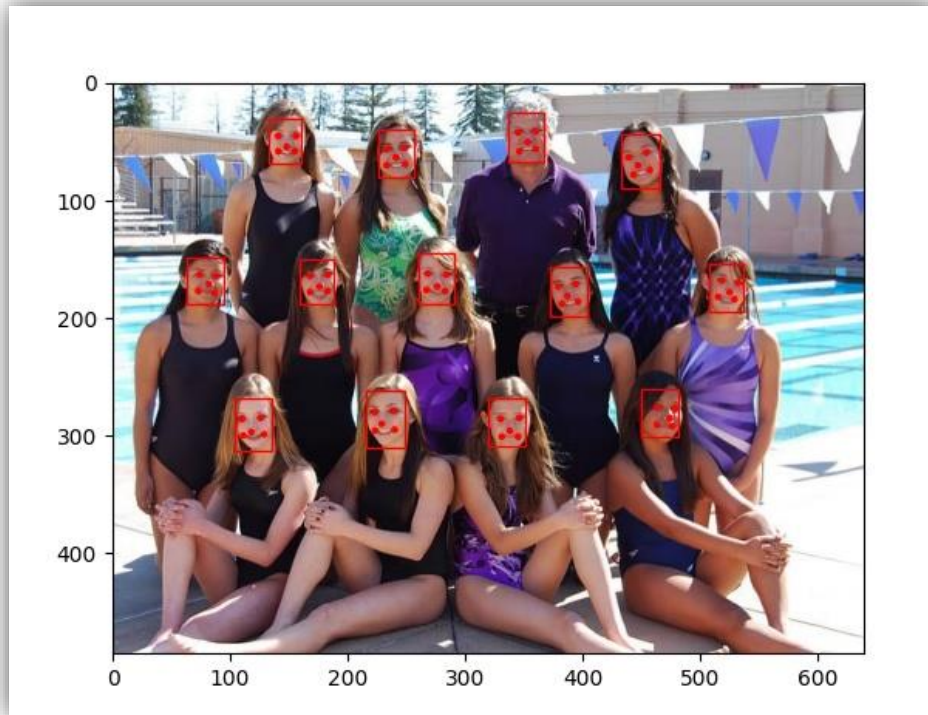
```

Συντάσσοντας το παρακάτω κομμάτι κώδικα:

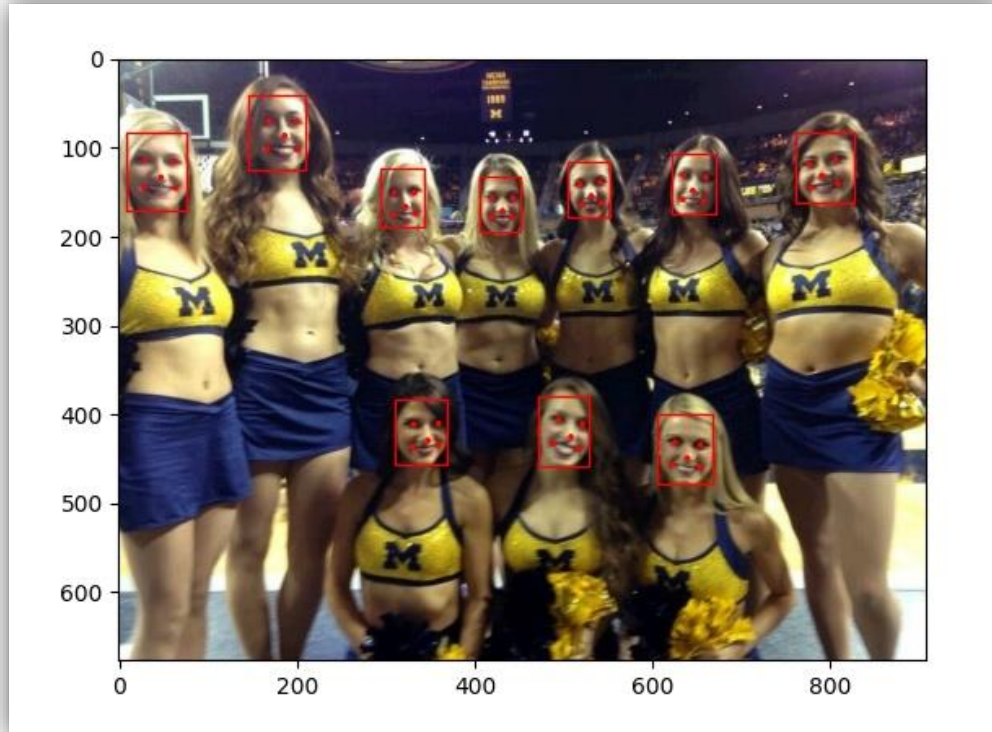
Είμαστε σε θέση για τις ίδιες εικόνες όπως και πριν να παρατηρήσουμε τα σαφώς βελτιωμένα αποτελέσματα, ενώ ταυτόχρονα να δούμε και την landmark detection διαδικασία.



Εικόνα 68: Αποτελέσματα Εικόνας 1



Εικόνα 69: Αποτελέσματα Εικόνας 2



Εικόνα 70: Αποτελέσματα Εικόνας 3

Παρατηρούμε λοιπόν, ότι ενώ η επιτυχία του κατηγορητητή είναι σαφώς βελτιωμένη σε σχέση με την feature based μέθοδο, η απόδοσή του, δηλαδή η ταχύτητα παραγωγής των αποτελεσμάτων είναι σαφώς χαμηλότερη, γεγονός που σε πραγματικές συνθήκες θα απαιτούσε σημαντικούς υπολογιστικούς πόρους.

Επίλογος

Η αναγνώριση προσώπων σε μη ελεγχόμενα περιβάλλοντα αποτελεί μία πολύπλοκη διαδικασία με αρκετές προτεινόμενες λύσεις. Ορισμένες από αυτές, βασίζονται σε συγκεκριμένες υποθέσεις όπως τα χαρακτηριστικά του προσώπου και προχωρούν σε υλοποιήσεις που προσφέρουν αξιόπιστα αποτελέσματα με σημαντική απόδοση, σε συστήματα πραγματικού χρόνου. Από την άλλη, στο συγκεκριμένο πεδίο εφαρμογών, η μηχανική μάθηση και περισσότερο η βαθιά μηχανική μάθηση με την χρήση μίας πληθώρας αρχιτεκτονικών είναι σε θέση να προσφέρει αξιόπιστες λύσεις παρόλο που ελλοχεύει ο κίνδυνος ύπαρξης σημαντικού υπολογιστικού κόστους. Στη συγκεκριμένη εργασία, πραγματοποιήθηκε μία σύνοψη ορισμένων βασικών εννοιών της μηχανικής μάθησης και διερευνήθηκε η θεωρία και τα αποτελέσματα σημαντικών αλγορίθμων στον τομέα της μηχανικής υπολογιστικής όρασης και ειδικότερα της αναγνώρισης προσώπου σε μη ελεγχόμενα περιβάλλοντα.

Βιβλιογραφία

- Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2010). Face Recognition by Independent Component Analysis. *IEEE Trans Neural Network*.
- Bhattacharyya, S. K., & Rahul, K. (2013). Face recognition by linear discriminant analysis.
- Cotter, N. (1990). The Stone-Weierstrass theorem and its application to neural networks. N.E. Cotter. *IEEE Transactions on Neural Networks*.
- Davies, E. (2005). The Importance of the Nearest Neighbor Classifier. *Statistical Pattern Recognition*.
- Dev, V. A., & Eden, M. R. (2019). Decision Trees. *Proceedings of the 9th International Conference on Foundations of Computer-Aided Process Design*.
- Gardner, M., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*.
- Heung-IISuk. (2017). An Introduction to Neural Networks and Deep Learning. Στο Heung-IISuk, *Deep Learning for Medical Image Analysis* (σσ. 3-24).
- Hoffman, J. I. (2015). Logistic Regression Analysis. *Biostatistics for Medical and Biomedical Practitioners*.
- Karamizadeh, S., & Abdullah, S. (2013). An Overview of Holistic Face Recognition.
- Kulkarni, A., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. Στο *Data Democracy*.
- Naive Bayes Classifier. (2018). *Digital Communications and Networks*.
- Paul, L. C., & Suman, A. (2012). Face recognition using principal component analysis method.
- Pisner, D. A., & Schnyer, D. M. (2020). Support Vector Machine. *Machine Learning*.
- Simon, R. (2013). Overfitting in prediction models – Is it a problem only in high dimensions? *Contemporary Clinical Trials*.

Simple Linear Regression. (2017). *Biostatistics and Computer-based Analysis of Health Data using SAS*.

Viola, P., & Jones, Michael. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001*.

Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks. Στο *Encyclopedia of Physical Science and Technology (Third Edition)*.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks. *IEEE Signal Processing Letters*.

Διαδικτυακές Πηγές

1. http://www.oramasolutions.com/?page_id=1866
2. <https://hayo.io/computer-vision/>
3. https://el.wikipedia.org/wiki/Τεχνητή_νοημοσύνη
4. https://el.wikipedia.org/wiki/Μηχανική_μάθηση
5. https://en.wikipedia.org/wiki/Deep_learning
6. <https://www.zendesk.com/blog/machine-learning-and-deep-learning/>
7. <https://www.mathworks.com/discovery/deep-learning.html>
8. <http://www.business-analytics.gr/news/1211-machine-learning-vs-statistics>
9. <http://aibook.csd.auth.gr/include/slides/Chap18.pdf>
10. http://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/8283/Alexandropoulou_Charikleia.pdf?sequence=1&isAllowed=y
11. https://repository.kallipos.gr/bitstream/11419/3382/1/02_chapter_04.pdf
12. https://el.wikipedia.org/wiki/Στατιστική_ταξινόμηση
13. https://el.wikipedia.org/wiki/Αναγνώριση_προτύπων