

Νευρωνικά δίκτυα

Αναπλ. Καθηγ. Στελιος Ζήμερας
Τμήμα Στατιστικής και Αναλογιστικών –
Χρηματοοικονομικών Μαθηματικών
Πανεπιστήμιο Αιγαίου
Σαμος

2021

Εισαγωγή

- Η ιδέα είναι να προσομοιώσουμε όσο το δυνατόν πιο στενά πώς λειτουργεί το ανθρώπινο νευρικό σύστημα - δηλαδή, την ικανότητά του να μαθαίνει και να διορθώνει τα λάθη. Αυτό είναι κύριο χαρακτηριστικό οποιοδήποτε νευρικό δίκτυο - είναι σε θέση να μαθαίνει και να ενεργεί ανεξάρτητα με βάση την προηγούμενη εμπειρία, κάθε φορά που κάνει λιγότερα λάθη.

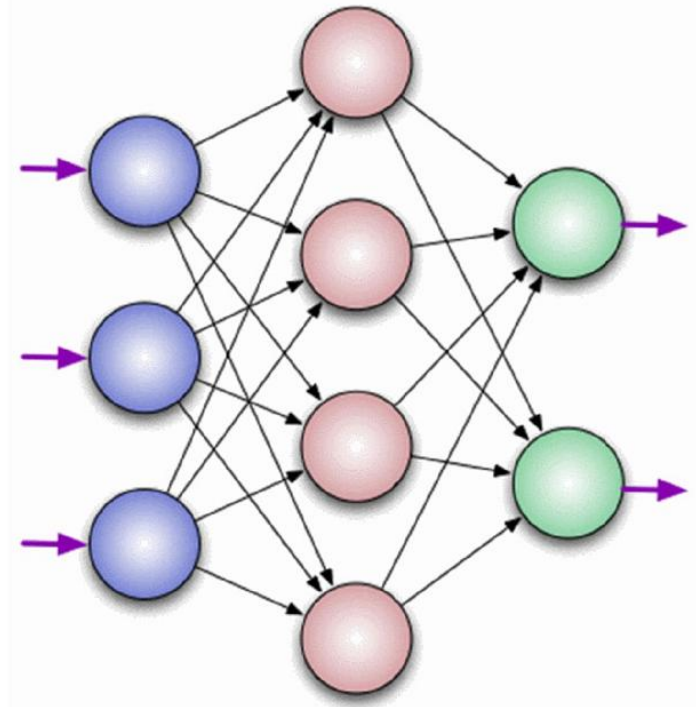
Εισαγωγή

- Το νευρικό δίκτυο μιμείται όχι μόνο τη δραστηριότητα, αλλά και τη δομή του ανθρώπινου νευρικού συστήματος. Ένα τέτοιο δίκτυο αποτελείται από μεγάλο αριθμό μεμονωμένων υπολογιστικών στοιχείων ("νευρώνες"). Στις περισσότερες περιπτώσεις, κάθε "νευρώνας" αναφέρεται σε ένα συγκεκριμένο στρώμα του δικτύου. Τα δεδομένα εισαγωγής επεξεργάζονται διαδοχικά σε όλα τα επίπεδα δικτύου. Οι παράμετροι κάθε «νευρώνα» μπορούν να αλλάξουν ανάλογα με τα αποτελέσματα που αποκτήθηκαν στα προηγούμενα σύνολα δεδομένων εισόδου, αλλάζοντας έτσι τη σειρά ολόκληρου του συστήματος.

Εισαγωγή

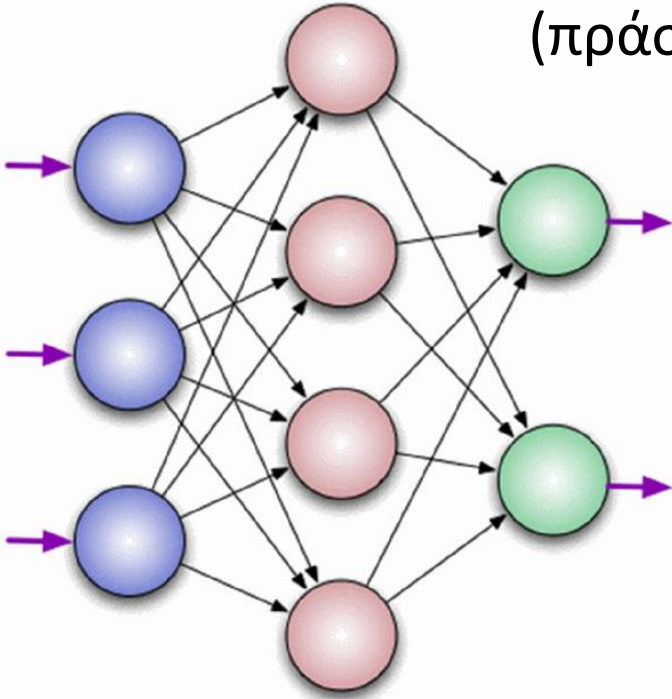
- Τα νευρικά δίκτυα χρησιμοποιούνται για την επίλυση σύνθετων προβλημάτων που απαιτούν αναλυτικούς υπολογισμούς παρόμοιους με αυτούς του ανθρώπινου εγκεφάλου.

Ο νευρώνας είναι μια υπολογιστική μονάδα που λαμβάνει πληροφορίες, εκτελεί απλούς υπολογισμούς σε αυτήν και τις μεταφέρει περαιτέρω. Χωρίζονται σε τρεις βασικούς τύπους:



Εισαγωγή

Χωρίζονται σε τρεις βασικούς τύπους: είσοδο (μπλε), κρυφό (κόκκινο) και έξοδο (πράσινο).



Στην περίπτωση που το νευρικό δίκτυο αποτελείται από έναν μεγάλο αριθμό νευρώνες, εισάγετε τον όρο στρώμα. Κατά συνέπεια, υπάρχει ένα επίπεδο εισόδου που λαμβάνει πληροφορίες, η κρυμμένα στρώματα (συνήθως όχι περισσότερο από 3) που το επεξεργάζονται και ένα επίπεδο εξόδου που εξάγει το αποτέλεσμα.

Κάθε ένας από τους νευρώνες έχει 2 κύριες παραμέτρους: δεδομένα εισόδου και δεδομένα εξόδου

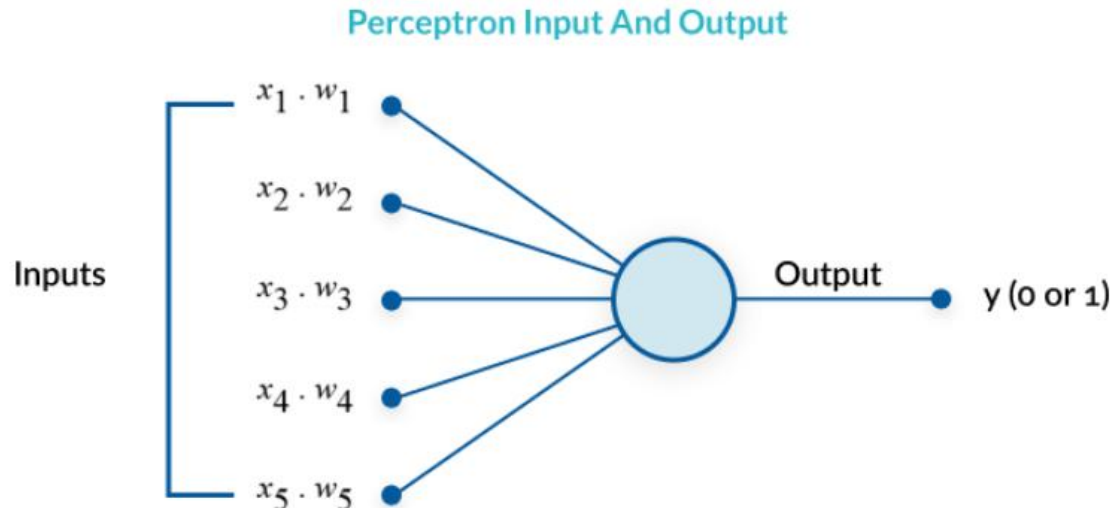
Εισαγωγή

- Οι συνολικές πληροφορίες όλων των νευρώνων από το προηγούμενο στρώμα εισέρχονται στο πεδίο εισόδου, μετά τον οποίο ομαλοποιείται χρησιμοποιώντας τη λειτουργία ενεργοποίησης (ως συνάρτηση $f(x)$) και μπαίνει στο πεδίο εξόδου.
- Το Νευρωνικό δίκτυο δομείται ως ένας κατευθυνόμενος γράφος με πολλούς κόμβους (στοιχεία επεξεργασίας) και τόξα (συνδέσμους) μεταξύ των κόμβων. Οι κόμβοι στο γράφο είναι σαν ανεξάρτητοι νευρώνες ενώ τα τόξα είναι σύνδεσμοι

Κατηγορίες

Perceptron

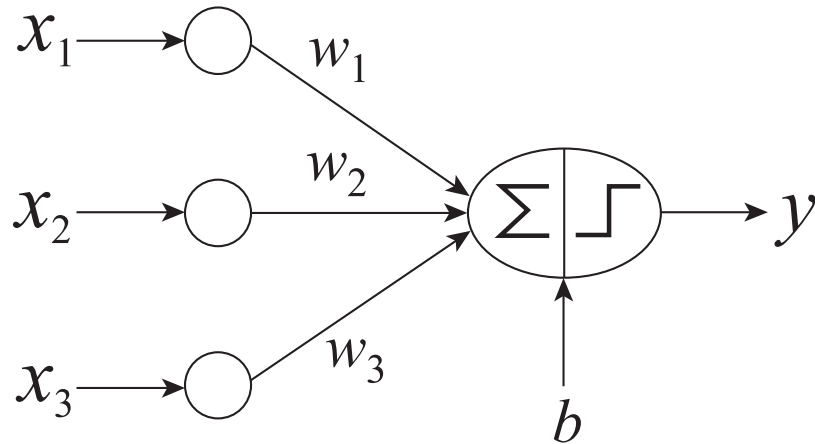
Το perceptron είναι ένας απλός αλγόριθμος δυαδικής (binary) ταξινόμησης. Στόχος του είναι να χωρίσει ένα σύνολο σημάτων εισόδου σε δύο μέρη - "ναι" και "όχι". Σε αντίθεση με πολλούς άλλους αλγορίθμους ταξινόμησης, το perceptron μοντελοποιήθηκε πάνω στη βασική μονάδα επεξεργασίας του ανθρώπινου εγκεφάλου, τον νευρώνα



Κατηγορίες

- Ένα perceptron είναι μια πολύ απλή μηχανή εκμάθησης. Μπορεί να δεχθεί μερικές εισόδους x_i , $i = 1, \dots, N$, κάθε μία από τις οποίες πολλαπλασιάζεται με ένα αντίστοιχο συντελεστή ή «βάρος» w_i , $i = 1, 2, \dots, N$ που υποδηλώνει πόσο σημαντική είναι η αντίστοιχη είσοδος, να αθροίσει τα επιμέρους γινόμενα ($x_i \cdot w_i$) και να παράγει μια έξοδο "0" ή "1" που αποτελεί και την απόφαση του perceptron για την κλάση (μία από τις δύο) που ανήκει η συγκεκριμένη είσοδος.

Κατηγορίες



$$y = \begin{cases} 1, & \text{if } \mathbf{w}^T \mathbf{x} + b > 0. \\ -1, & \text{otherwise.} \end{cases}$$

Οι τιμές εισόδου x_i μπορούν να είναι πραγματικές (+ ή -), δυαδικές (0,1) ή διπολικές (-1,+1).

$$net = x_1 w_1 + x_2 w_2 + x_3 w_3 = \sum_{i=1}^3 x_i w_i$$

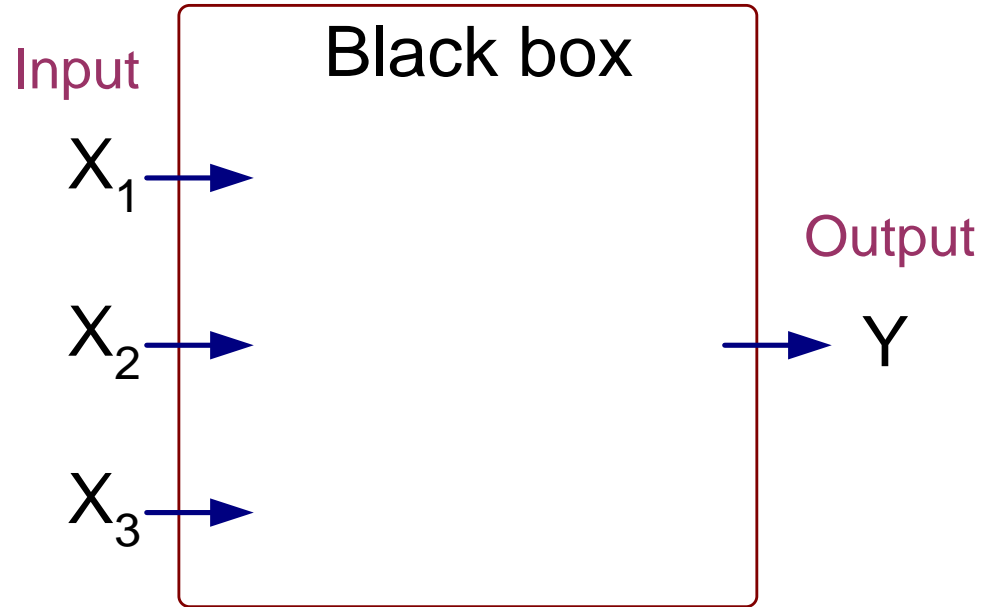
Έξοδος
Τεχνητού
Δικτύου

Νευρωνικού

$$y = f\left(\sum_{i=1}^n x_i w_i\right)$$

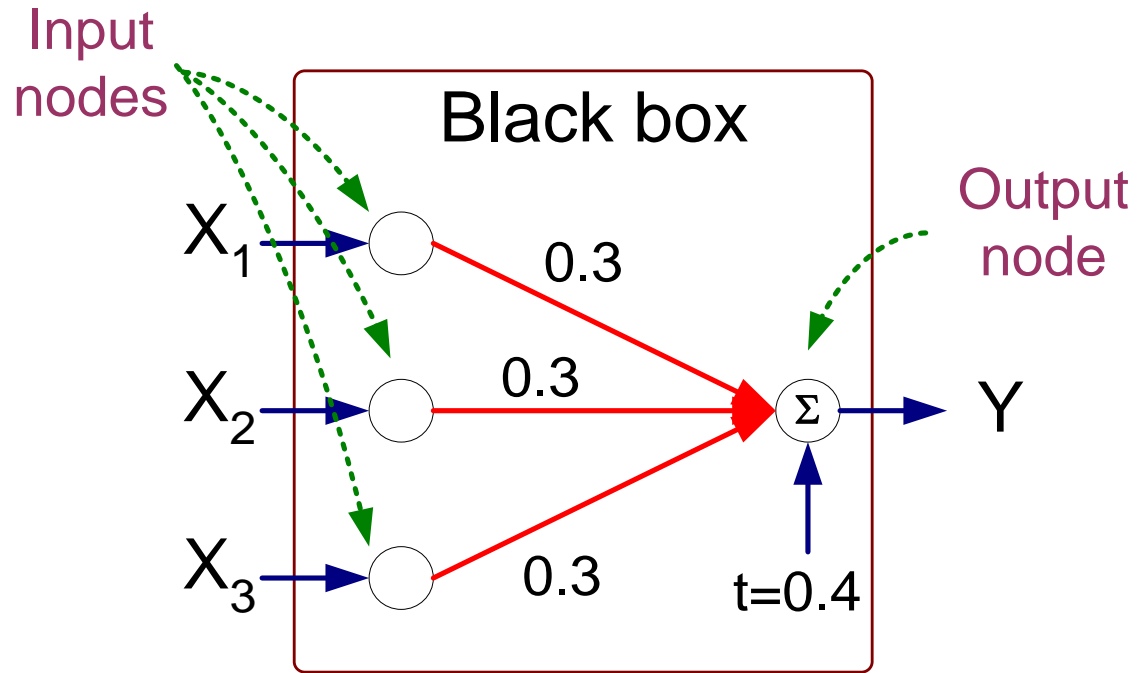
Παράδειγμα

X_1	X_2	X_3	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	-1
0	1	0	-1
0	1	1	1
0	0	0	-1



Παράδειγμα

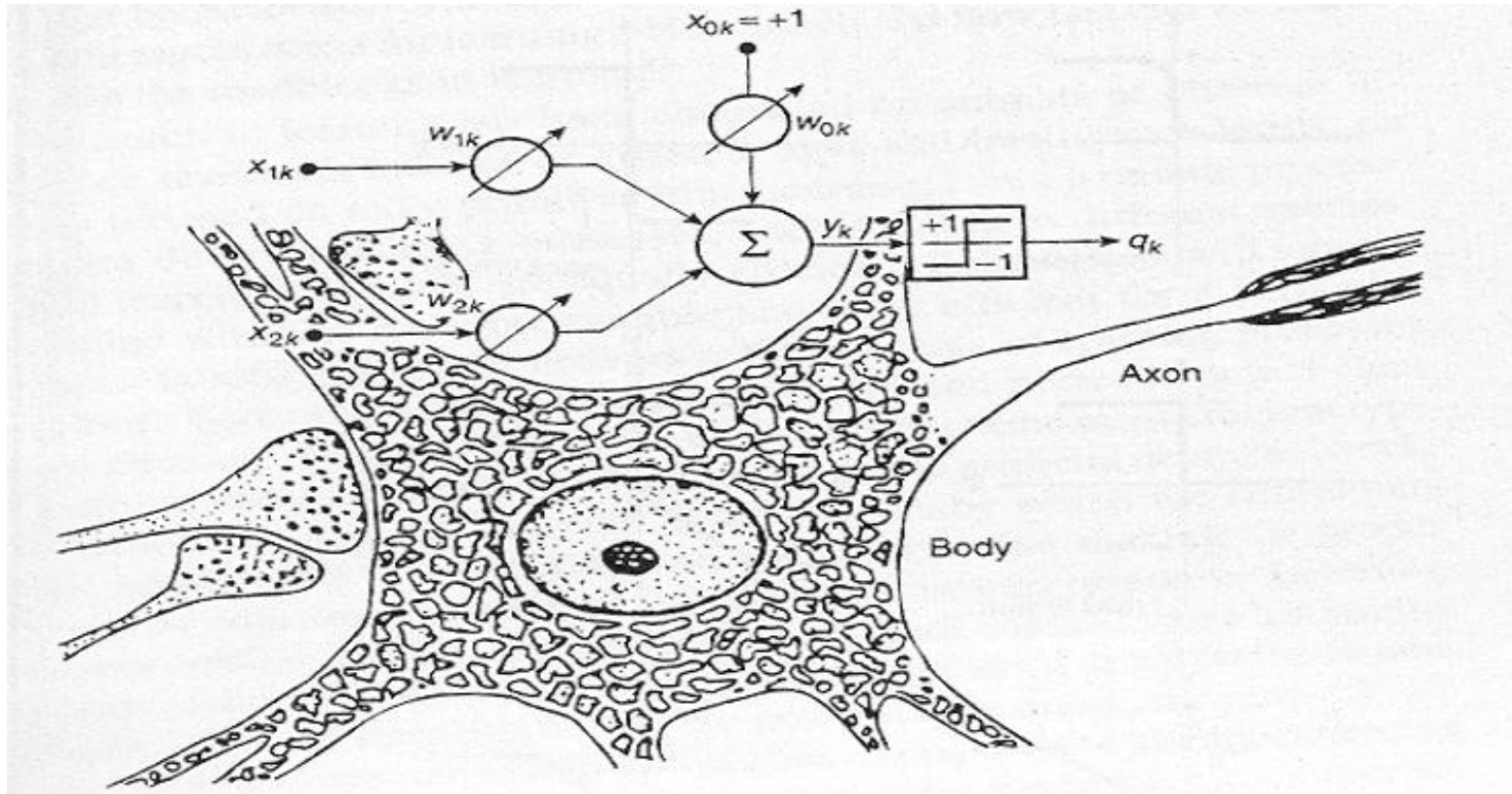
X_1	X_2	X_3	Y
1	0	0	-1
1	0	1	1
1	1	0	1
1	1	1	1
0	0	1	-1
0	1	0	-1
0	1	1	1
0	0	0	-1



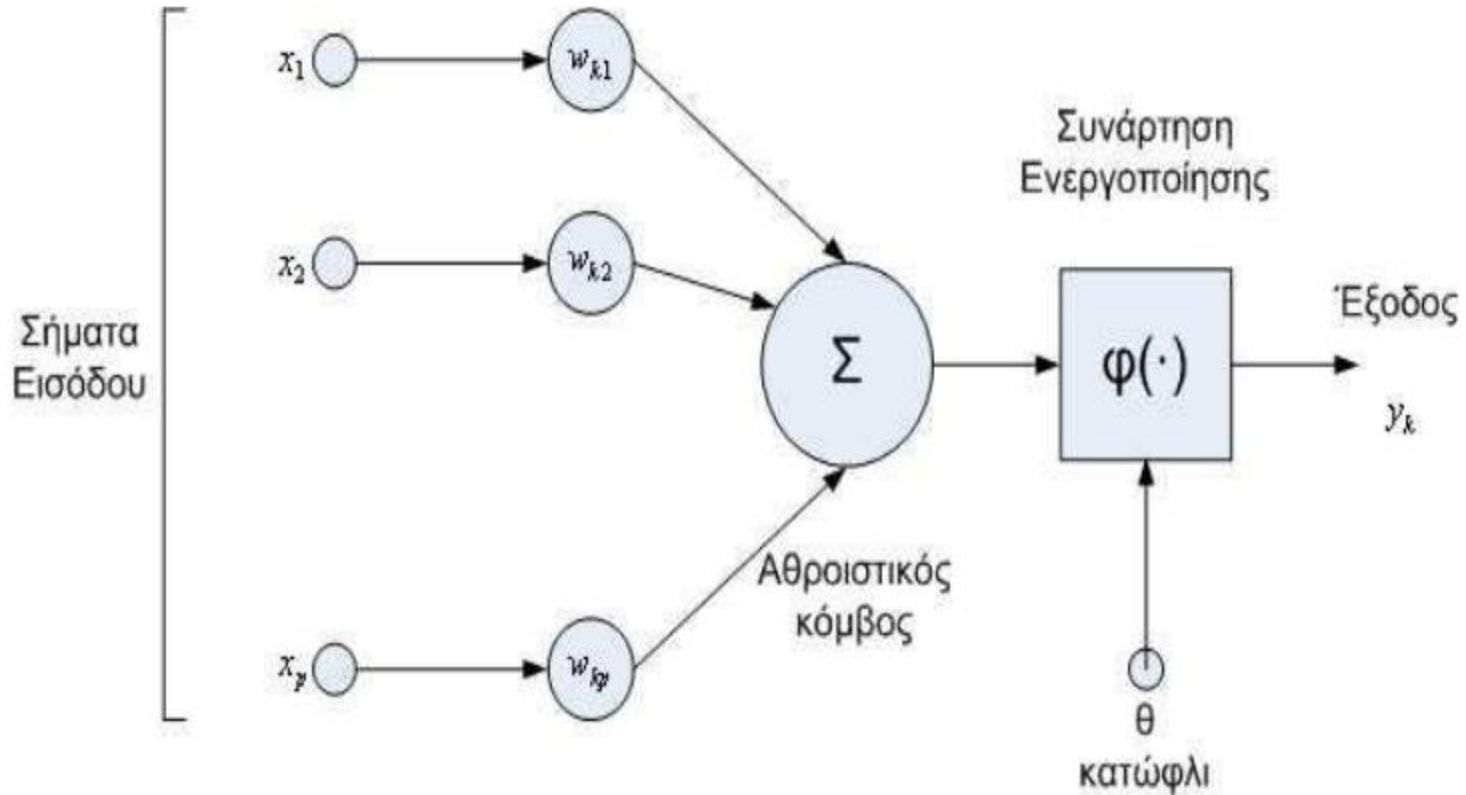
$$Y = \text{sign}(0.3X_1 + 0.3X_2 + 0.3X_3 - 0.4)$$

$$\text{where } \text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Παράδειγμα



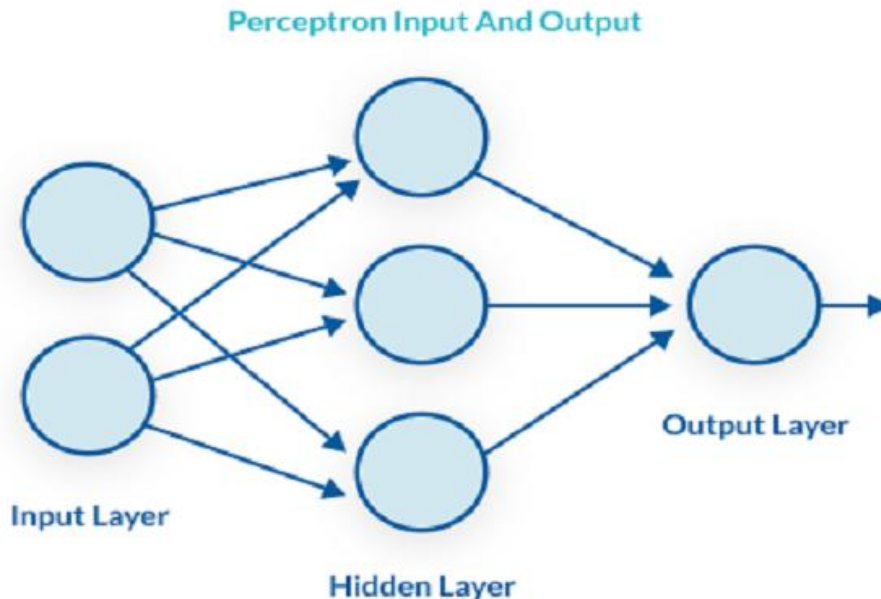
Γενική μορφή



Κατηγορίες

Multilayer Perceptron

Ένα perceptron πολλαπλών στρωμάτων (Multilayer Perceptron, MLP) είναι ένα perceptron που συνδυάζεται με άλλα πρόσθετα perceptrons, στοιβαγμένα σε επάλληλα στρώματα, για την επίλυση πιο σύνθετων προβλημάτων από την απλή δυαδική ταξινόμηση.



Κατηγορίες

- Κάθε perceptron δέχεται εισόδους στο πρώτο αριστερά στρώμα (το στρώμα εισόδου – input layer), τις πολλαπλασιάζει επί τα αντίστοιχα βάρη και στέλνει τα αποτελέσματα (τις εξόδους του) σε όλα τα perceptrons του δεύτερου στρώματος (το κρυμμένο στρώμα – hidden layer). Οι έξοδοι του δεύτερου στρώματος αφού έχουν πολλαπλασιαστεί με τα αντίστοιχα βάρη στέλνονται σε όλους τους κόμβους του τρίτου και τελικού στρώματος δεξιά (στρώμα εξόδου – output layer).
- Κάθε perceptron στέλνει σήματα εξόδου προς κάθε perceptron στο επόμενο στρώμα. Για κάθε σήμα, το perceptron χρησιμοποιεί διαφορετικά βάρη

Κατηγορίες

- Οι νευρώνες του κρυφού στρώματος δέχονται τις τιμές των νευρώνων εισόδου πολλαπλασιασμένες με τα βάρη των συνδέσεων, τις αθροίζουν και μετασχηματίζουν το άθροισμα σύμφωνα με τη συνάρτηση μετασχηματισμού. Οι κρυφοί νευρώνες είναι καθοριστικής σημασίας για την καταγραφή των σύνθετων σχέσεων των δεδομένων. Οι τιμές εξόδου των κρυφών νευρώνων, πολλαπλασιασμένες με τα βάρη των συνδέσεων, διαβιβάζονται στους νευρώνες του στρώματος εξόδου (output layer).

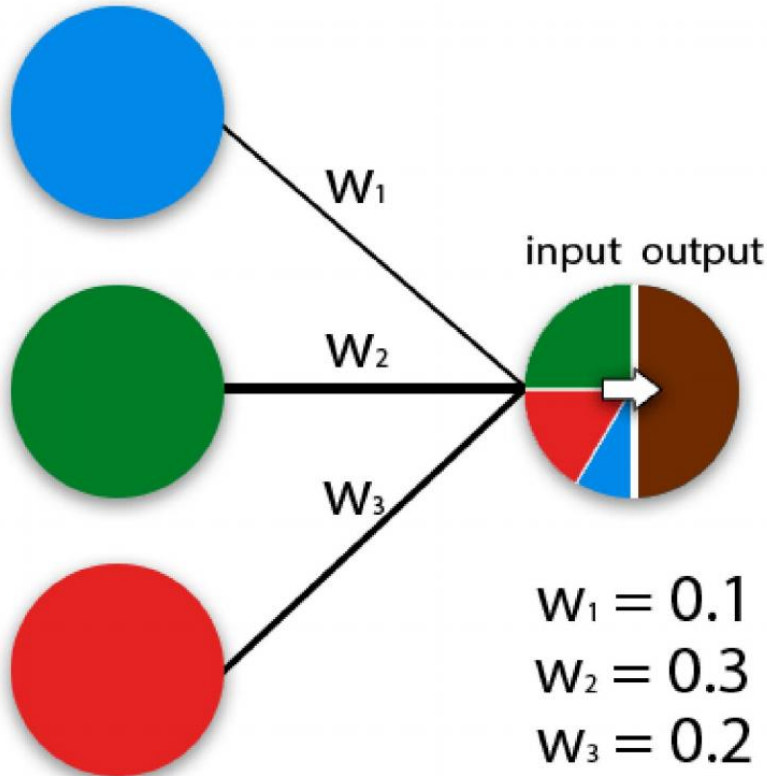
Κατηγορίες

- Κάθε στρώμα μπορεί να έχει μεγάλο αριθμό perceptrons, και επιπλέον μπορεί να υπάρχουν πολλαπλά στρώματα, οπότε το perceptron πολλαπλών στρωμάτων μπορεί να γίνει άμεσα ένα πολύ περίπλοκο σύστημα.
- Στο κλασικό perceptron, η λειτουργία απόφασης αποτελεί μια βαθμίδα και η έξοδος είναι δυαδική.
- Στα MLPs, μπορούν να χρησιμοποιηθούν και άλλες λειτουργίες ενεργοποίησης οι οποίες έχουν ως αποτέλεσμα εξόδους πραγματικών τιμών, συνήθως μεταξύ 0 και 1 ή μεταξύ -1 και 1.

Σύναψη

Η σύναψη είναι μια σύνδεση μεταξύ δύο νευρώνων. Οι συνάψεις έχουν 1 παράμετρο - βάρος. Χάρη σε αυτόν, οι πληροφορίες εισαγωγής αλλάζουν όταν μεταδίδονται από τον ένα νευρώνα στον άλλο.

Ας πούμε ότι υπάρχουν 3 νευρώνες που μεταδίδουν πληροφορίες στον επόμενο. Τότε έχουμε 3 βάρη που αντιστοιχούν σε καθέναν από αυτούς τους νευρώνες. Για τον νευρώνα με το μεγαλύτερο βάρος, αυτές οι πληροφορίες θα κυριαρχούν στον επόμενο νευρώνα (για παράδειγμα, ανάμειξη χρώματος).



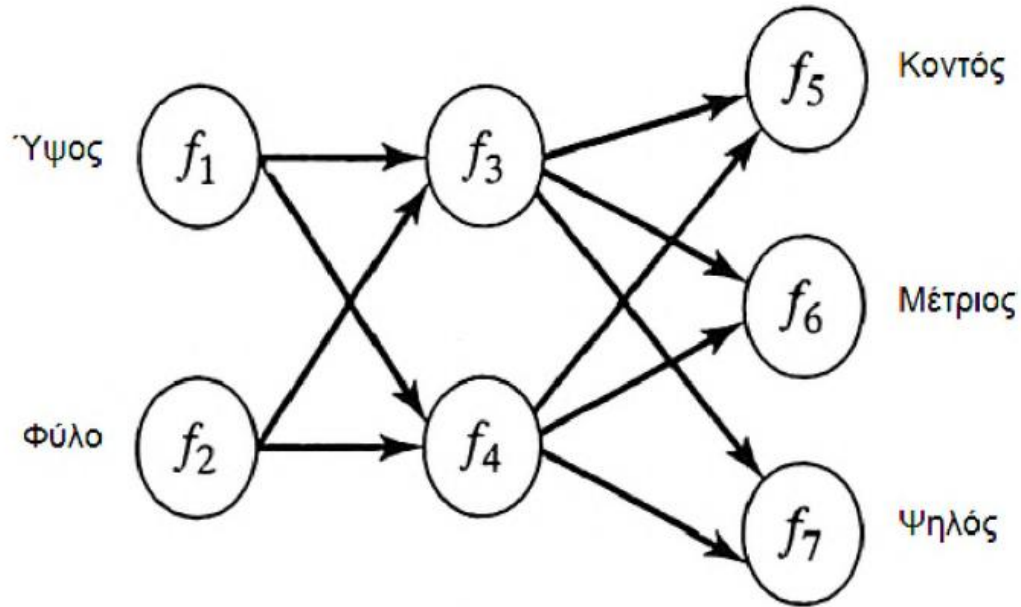
Λειτουργία νευρωνικού δικτύου

Το νευρωνικό δίκτυο μπορεί να θεωρηθεί ως ένας κατευθυνόμενος γράφος με πηγή (είσοδο), καταβόθρα (έξοδο) και εσωτερικούς (κρυμμένους) κόμβους. Για να εκτελεστεί η λειτουργία της εξόρυξης γνώσης μέσω NN, μία εγγραφή δίνεται στην είσοδο, μέσω των κόμβων εισόδου, και ο κόμβος εξόδου καθορίζει ποια είναι η πρόβλεψη. Αντίθετα με τα δένδρα απόφασης, τα οποία διαθέτουν μόνο ένα κόμβο εισόδου, την ρίζα του δένδρου, το NN έχει έναν κόμβο εισόδου για κάθε τιμή γνωρίσματος το οποίο θα εξεταστεί για την επίλυση του προβλήματος εξόρυξης γνώσης που έχει τεθεί.

Παράδειγμα

Όνομα	Φύλο	Ύψος	1 ^η Κατηγοριοποίηση	2 ^η Κατηγοριοποίηση
Χριστίνα	Θ	1,6	Κοντός	Μέτριος
Γιάννης	A	2	Ψηλός	Μέτριος
Μαρία	Θ	1,9	Μέτριος	Ψηλός
Μάρθα	Θ	1,88	Μέτριος	Ψηλός
Στεφανία	Θ	1,7	Κοντός	Μέτριος
Βασίλης	A	1,85	Μέτριος	Μέτριος
Κατερίνα	Θ	1,6	Κοντός	Μέτριος
Ντίνος	A	1,7	Κοντός	Μέτριος
Σπύρος	A	2,2	Ψηλός	Ψηλός
Στέργιος	A	2,1	Ψηλός	Ψηλός
Σοφία	Θ	1,8	Μέτριος	Μέτριος
Τάκης	A	1,95	Μέτριος	Μέτριος
Στέλλα	Θ	1,9	Μέτριος	Ψηλός
Αθηνά	Θ	1,8	Μέτριος	Μέτριος
Γιάννα	Θ	1,75	Μέτριος	Μέτριος

Λειτουργία νευρωνικού δικτύου



Νευρωνικό δίκτυο όπου δέχεται ως είσοδο την συνάρτηση των γνωρισμάτων του ύψους και του φύλου και έξοδο τις συναρτήσεις των κατηγοριών κοντός, μέτριος και ψηλός. Οι συναρτήσεις f_3 και f_4 είναι οι εσωτερικοί κόμβοι τους οποίους αυθαίρετα καθορίσαμε σε δύο

Λειτουργία νευρωνικού δικτύου

Ορισμός Ένα νευρωνικό δίκτυο είναι ένας κατευθυνόμενος γράφος $F = \langle V, A \rangle$ με κορυφές $V = \{1, 2, \dots, n\}$ και τόξα $A = \{ \langle i, j \rangle \mid 1 \leq i, j \leq n \}$, με τους παρακάτω περιορισμούς:

- Το V χωρίζεται σε ένα σύνολο από κόμβους εισόδου V_i , κρυμμένους κόμβους V_h και κόμβους εξόδου V_o .
- Οι κορυφές χωρίζονται σε στρώματα $(1, 2, \dots, k)$ με όλους τους κόμβους εισόδου στο πρώτο στρώμα, όλους τους κρυμμένους κόμβους στα στρώματα $2, \dots, k-1$ (κρυμμένα στρώματα) και όλους τους κόμβους εξόδους στο στρώμα k .
- Κάθε τόξο $\langle i, j \rangle$ πρέπει να έχει τον κόμβο i στο στρώμα $h-1$ και τον κόμβο j στο στρώμα h .
- Το τόξο $\langle i, j \rangle$ έχει σαν ετικέτα την αριθμητική τιμή w_{ij} .
- Ο κόμβος i έχει για ετικέτα τη συνάρτηση f_i .

Λειτουργία νευρωνικού δικτύου

Ορισμός: Ένα μοντέλο νευρωνικού δικτύου είναι ένα υπολογιστικό μοντέλο που αποτελείται από τρία μέρη:

- Το γράφο του νευρωνικού δικτύου, ο οποίος ορίζει τη δομή δεδομένων
- Τον αλγόριθμο μάθησης που δείχνει με ποιον τρόπο γίνεται η μάθηση
- Τις τεχνικές ανάκλησης που καθορίζουν πώς η πληροφορία παρέχεται από το δίκτυο

Λειτουργία νευρωνικού δικτύου

- Η έξοδος κάθε κόμβου βασίζεται στην συνάρτηση ενεργοποίησης (activation function). Η συνάρτηση εφαρμόζεται σε σύνολο των τιμών των εισόδων που έρχονται μέσω των τόξων εισόδου λαμβάνοντας υπόψη και τα αντίστοιχα βάρη.
- Μια συνάρτηση ενεργοποίησης καλείται και κανόνας διέγερσης παραπέμποντας στη λειτουργία του ανθρώπινου εγκεφάλου. Όταν η είσοδος σε έναν νευρώνα είναι μεγάλη, τότε αυτός διεγείρεται στέλνοντας ένα ηλεκτρικό σήμα στο νευρίτη (έξοδος). Τα νευρωνικά δίκτυα, μιμούνται την παραπάνω διαδικασία παράγοντας έξοδο σε έναν κόμβο μόνο όταν η συνάρτηση ενεργοποίησης δίνει μια αρκετά μεγάλη τιμή.

Λειτουργία νευρωνικού δικτύου

- Στο στάδιο αυτό, μετασχηματίζεται το άθροισμα των τιμών εισόδου, με χρήση μιας συνάρτησης γνωστής ως συνάρτηση ενεργοποίησης (activation function) ή συνάρτηση μετασχηματισμού. Η τιμή που υπολογίζεται είναι η τιμή εξόδου του νευρώνα. Διάφορες συναρτήσεις μπορούν να χρησιμοποιηθούν ως συναρτήσεις ενεργοποίησης. Τέτοιες συναρτήσεις είναι η συνάρτηση ημιτόνου, η συνάρτηση συνημίτονου, η συνάρτηση υπερβολικής εφαπτομένης κλπ. Συνήθως όμως χρησιμοποιείται η Σιγμοειδής συνάρτηση, επειδή είναι απλή και μη γραμμική και επειδή μοιάζει με τη συμπεριφορά των πραγματικών νευρώνων.

Λειτουργία νευρωνικού δικτύου

Η επίλυση των προβλημάτων κατηγοριοποίησης χρησιμοποιώντας νευρωνικά δίκτυα περιλαμβάνει τα εξής βήματα:

- **Καθορισμός του αριθμού των εξόδων και των χαρακτηριστικών που θα χρησιμοποιηθούν σαν είσοδοι.** Καθορισμός του αριθμού των κρυμμένων στρωμάτων. Το βήμα αυτό εκτελείται από τον ειδικό του τομέα.
- **Καθορισμός των βαρών και των συναρτήσεων που θα χρησιμοποιηθούν.**
- **Μετάδοση της κάθε πλειάδας του συνόλου εκπαίδευσης μέσω του δικτύου και αξιολόγηση της εξόδου ως προς το πραγματικό αποτέλεσμα.** Αν η κατηγοριοποίηση είναι ακριβής, οι ετικέτες προσαρμόζονται κατάλληλα ώστε να σιγουρευτεί το ότι αυτή η κατηγοριοποίηση έχει υψηλότερο βάρος για την έξοδο την επόμενη φορά. Από την άλλη, αν η κατηγοριοποίηση δεν είναι σωστή, τα βάρη πρέπει να προσαρμοστούν ώστε να δώσουν μια χαμηλότερη τιμή για αυτή την κατηγορία.

Λειτουργία νευρωνικού δικτύου

- Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδο τους με το αντίστοιχο συνοπτικό βάρος και υπολογίζουν το ολικό άθροισμα. Έτσι το άθροισμα τροφοδοτεί τη συνάρτηση ενεργοποίησης, την οποία υλοποιεί κάθε κόμβος. Κάθε φορά λαμβάνεται η τιμή της συνάρτησης και η έξοδος του νευρώνα για τις τρέχουσες τιμές.

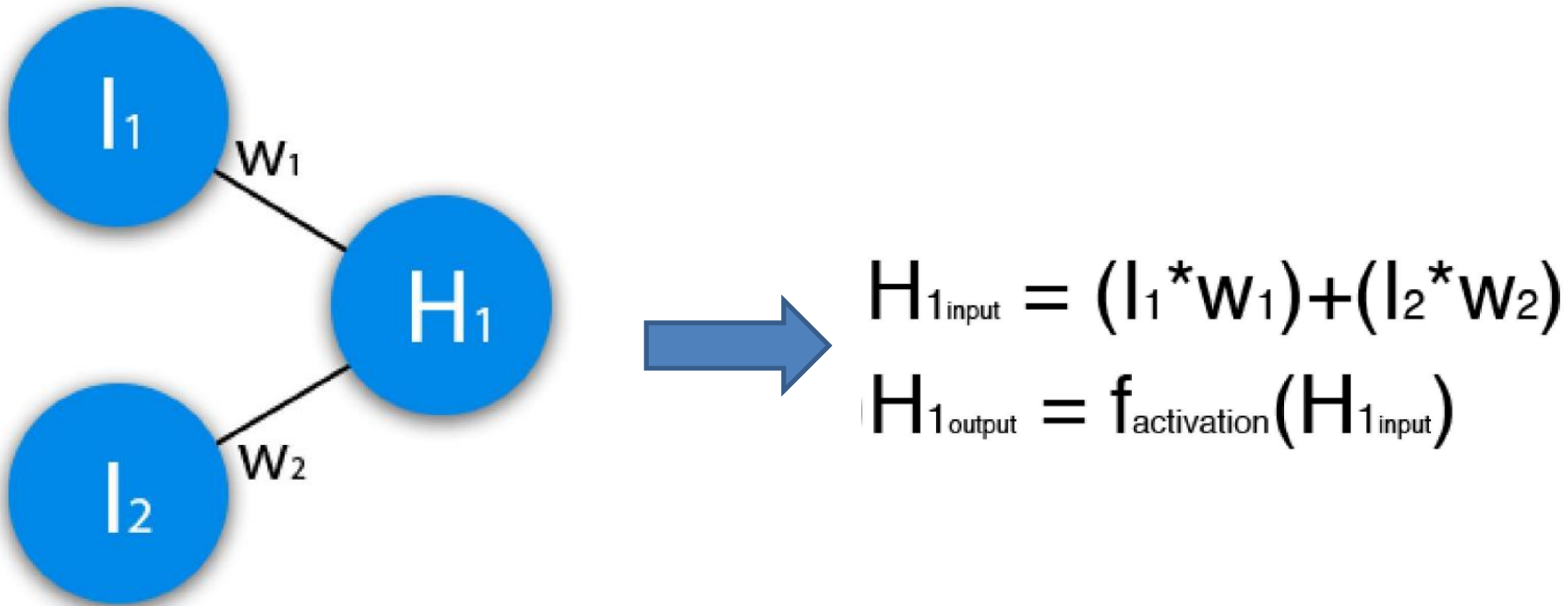
$$y_k = \varphi\left(\sum_{i=0}^N x_{ki} w_{ki}\right),$$

Λειτουργία νευρωνικού δικτύου

$$y_k = \varphi\left(\sum_{i=0}^N x_{ki} w_{ki}\right),$$

- Όπου x_{ki} είναι η i -οστή εισοδος του k νευρώνα, w_{ki} το i -οστό συνοπτικό βάρος του k νευρώνα, φ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου και y_k η έξοδος του k νευρώνα.
- Στον k οστό νευρώνα έχουμε το συνοπτικό βάρος w_{ki} όπου καλείται πόλωση ή κατώφλι. Αν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την αρχική του τιμή που θα πάρει και θα είναι 1, τότε ο νευρώνας ενεργοποιείται. Αν είναι μικρότερο από την τιμή της εισόδου, τότε ο νευρώνας παραμένει ανενεργός.

Λειτουργία νευρωνικού δικτύου



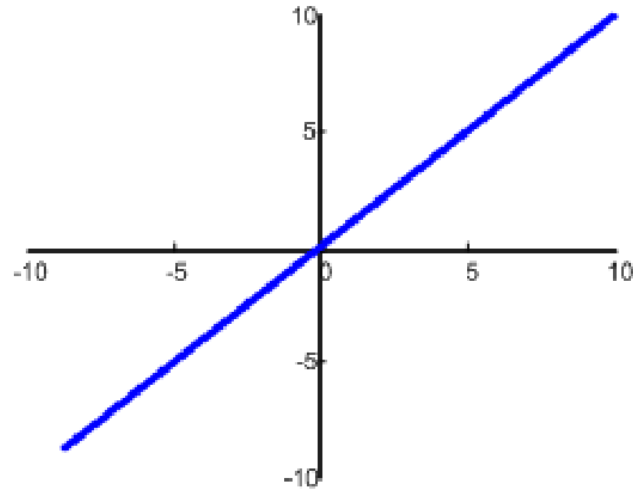
τα γράμματα δηλώνουν τους εισερχόμενους νευρώνες, το γράμμα H - ο κρυμμένος νευρώνας και το γράμμα w - τα βάρη.

Ο τύπος δείχνει ότι οι πληροφορίες εισαγωγής είναι το άθροισμα όλων των δεδομένων εισόδου πολλαπλασιασμένων με τα αντίστοιχα βάρη.

Λειτουργία νευρωνικού δικτύου

- Μια λειτουργία ενεργοποίησης είναι ένας τρόπος για να ομαλοποιήσετε τα δεδομένα εισόδου
- Γραμμική συνάρτηση

$$f(x) = x$$



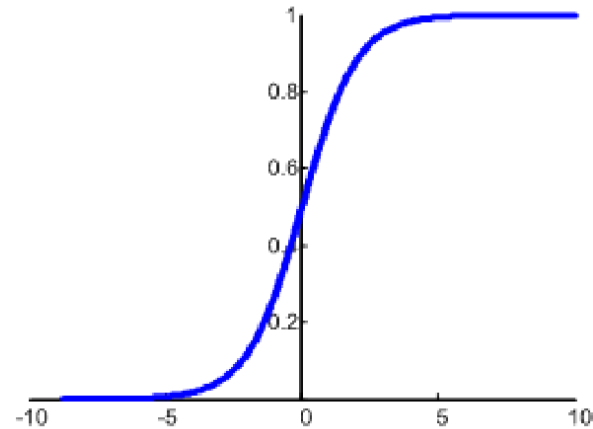
Αυτή η λειτουργία σχεδόν ποτέ δεν χρησιμοποιείται, εκτός εάν χρειάζεται να δοκιμάσετε ένα νευρωνικό δίκτυο ή να μεταφέρετε μια τιμή χωρίς μετασχηματισμούς.

Λειτουργία νευρωνικού δικτύου

- Σιγμοειδές

$$f(x) = \frac{1}{1+e^{-x}}$$

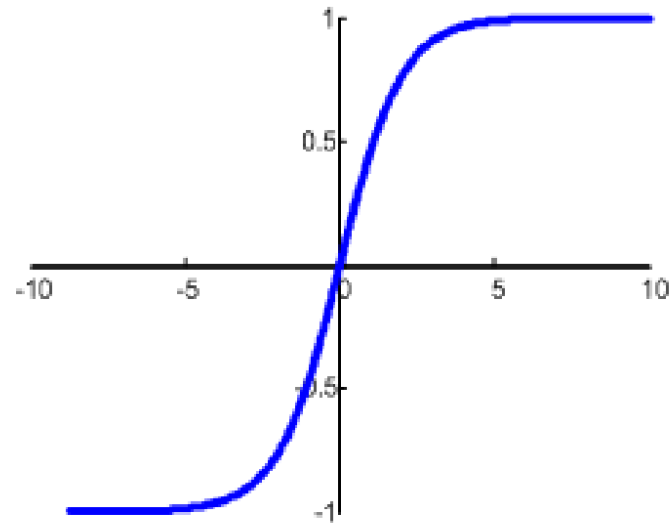
Αυτή είναι η πιο κοινή συνάρτηση ενεργοποίησης και το εύρος τιμών της.



- Υπερβολική εφαπτομένη

$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Είναι λογικό να χρησιμοποιείτε την υπερβολική εφαπτομένη μόνο όταν οι τιμές σας μπορούν να είναι τόσο αρνητικές όσο και θετικές, καθώς το εύρος της συνάρτησης είναι $[-1,1]$.



Εκπαίδευση Δικτύου

- Η εκπαίδευση ενός δικτύου συνίσταται στη ρύθμιση των βαρών των συνδέσεων και απαιτείται ένα σύνολο εκπαίδευσης και ένα σύνολο ελέγχου. Το σύνολο εκπαίδευσης χρησιμοποιείται για τον καθορισμό των βαρών των συνδέσεων και το σύνολο ελέγχου χρησιμοποιείται για την εκτίμηση της επίδοσης του μοντέλου. Στην επιβλεπόμενη μάθηση, κάθε παρατήρηση του συνόλου εκπαίδευσης περιλαμβάνει και την τιμή της κλάσης της συγκεκριμένης παρατήρησης.

Αλγόριθμος Αντίστροφης Μετάδοσης Σφάλματος

- Για να εκπαιδευση το δίκτυο, ο αλγόριθμος της Αντίστροφης Μετάδοσης Σφάλματος (Backpropagation) εφαρμόζει μια επαναληπτική διαδικασία όπου τα δεδομένα εκπαίδευσης εφαρμόζονται στο δίκτυο και στη συνέχεια υπολογίζεται στην έξοδο μια πρόβλεψη για την κλάση της παρατήρησης.
- Η πρόβλεψη αυτή συγκρίνεται με την πραγματική κλάση και αμέσως μετά αλλάζουν τα βάρη των συνδέσεων ώστε να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα μεταξύ των πρόβλεψής και της πραγματικής τιμής. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να ικανοποιηθούν οι συνθήκες τερματισμού.

Αλγόριθμος Αντίστροφης Μετάδοσης Σφάλματος

Το σφάλμα υπολογίζεται από την τετραγωνική συνάρτηση σφάλματος E :

$$\tilde{E} = \sum (t-o)^2$$

όπου:

o : Η πραγματική έξοδος του νευρωνίου του στρώματος εξόδου.

t : Η επιθυμητή έξοδος του νευρωνίου για το ορθό αποτέλεσμα.

Προβληματικές

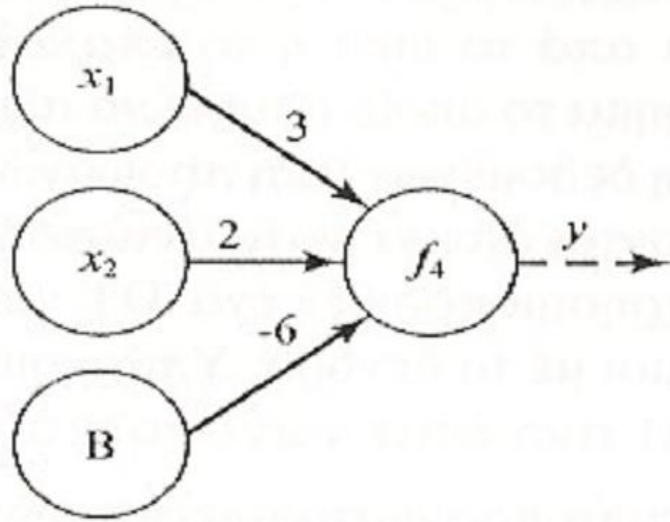
Θα πρέπει να μελετηθούν τα εξής θέματα

- **Χαρακτηριστικά (αριθμός κόμβων εισόδου):** Το θέμα αυτό είναι παρόμοιο με την επιλογή χαρακτηριστικών διάσπασης στα δένδρα απόφασης.
- **Αριθμός κρυμμένων στρωμάτων:** Στην πιο απλή περίπτωση υπάρχει μόνο ένα κρυμμένο στρώμα.
- **Αριθμός κρυμμένων κόμβων:** Αυτό είναι ένα αρκετά δύσκολο ζήτημα. Έχουν γίνει πολλές μελέτες που προσπαθούν να απαντήσουν σε αυτή την ερώτηση. Αυτό που είναι σίγουρο είναι ότι ο αριθμός των κρυμμένων κόμβων εξαρτάτε από τη δομή του νευρωνικού δικτύου, το είδος των συναρτήσεων ενεργοποίησης, τον αλγόριθμο εκπαίδευσης και το πρόβλημα που λύνεται. Αν αυτός ο αριθμός είναι μικρός, έχουμε αυξημένες πιθανότητες να συναντήσουμε το φαινόμενο της υποπροσαρμογής και η συνάρτηση να μη μαθευτεί. Αντίθετα, αν χρησιμοποιήσουμε πολλούς κόμβους έχουμε αντίστοιχες πιθανότητες να συναντήσουμε υπερπροσαρμογή. Οι εμπειρικοί κανόνες στηρίζονται στο μέγεθος του συνόλου εκπαίδευσης.
- **Δεδομένα εκπαίδευσης:** Αν έχουμε πολλά δεδομένα εκπαίδευσης, τότε μπορεί το μοντέλο να αντιμετωπίζει προβλήματα υπερπροσαρμογής ενώ με λίγα δεδομένα μπορεί να μην είναι ικανό να εκτελεί την κατηγοριοποίηση με μεγάλη ακρίβεια.

Προβληματικές

- **Αριθμός εξόδων:** Συνήθως έχουμε τόσους κόμβους εξόδου όσες είναι οι κατηγορίες. Ωστόσο αυτό δε γίνεται πάντα στην πράξη. π.χ. αν έχουμε δύο κατηγορίες, μπορούμε να έχουμε μόνο έναν κόμβο με την τιμή του αποτελέσματος να δείχνει την πιθανότητα της αντίστοιχης κατηγορίας. Εύκολα μπορούμε να υπολογίσουμε την πιθανότητα της άλλης κατηγορίας.
- **Διασυνδέσεις:** Στην πιο απλή περίπτωση κάθε κόμβος συνδέεται με όλους τους κόμβους του επόμενου επιπέδου.
- **Βάρη:** Το βάρος ενός τόξου δείχνει το σχετικό βάρος μεταξύ των κόμβων που συνδέει. Τα αρχικά βάρη είναι μικροί, θετικοί και τυχαίοι αριθμοί.
- **Συναρτήσεις ενεργοποίησης:** Μπορούν να χρησιμοποιηθούν διάφορες συναρτήσεις.
- **Τεχνική μάθησης:** Η τεχνική που χρησιμοποιείται για τον καθορισμό των βαρών λέγεται τεχνική μάθησης. Η πιο γνωστές τεχνικές μάθησης είναι κάποια μορφή οπισθοδρόμησης (backpropagation). Η οπισθοδρόμηση είναι μια τεχνική μάθησης που προσαρμόζει τα βάρη με το να μεταδίδει τις αλλαγές προς τα πίσω, δηλαδή από τους κόμβους εξόδου στους κόμβους εισόδου.
- **Τερματισμός:** Η μάθηση τερματίζει είτε όταν τελειώσουν τα δεδομένα εκπαίδευσης είτε χρησιμοποιώντας κάποιες ενδείξεις του χρόνου ή του ρυθμού εμφάνισης λαθών.

Παράδειγμα



Τα βάρη είναι 3, 2 και -6 αντίστοιχα.

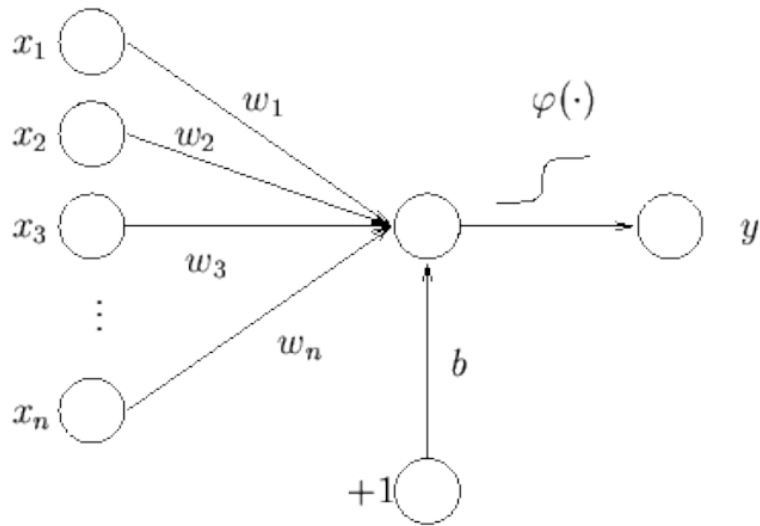
Η συνάρτηση ενεργοποίησης είναι f_4 και είναι αυτή που εφαρμόζεται στην τιμή

$$S = 3x_1 + 2x_2 - 6.$$

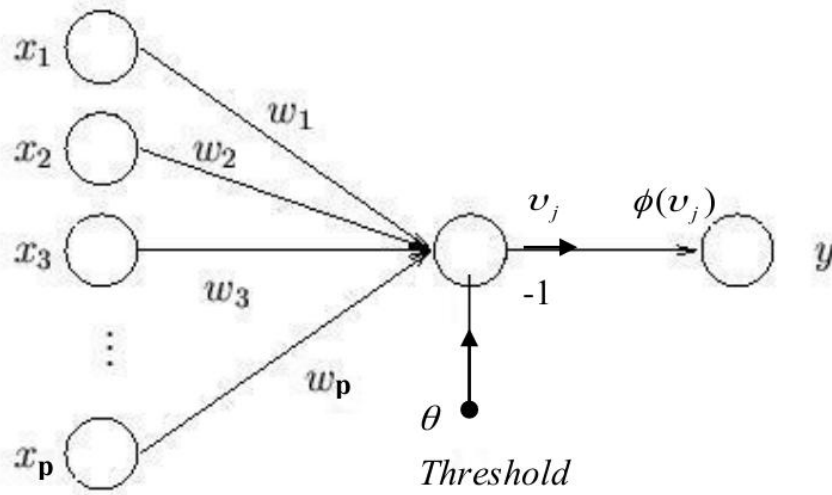
συνάρτηση ενεργοποίησης

$$f_4 = \begin{cases} 1, S > 0 \\ 0, \text{αλλιώς} \end{cases}$$

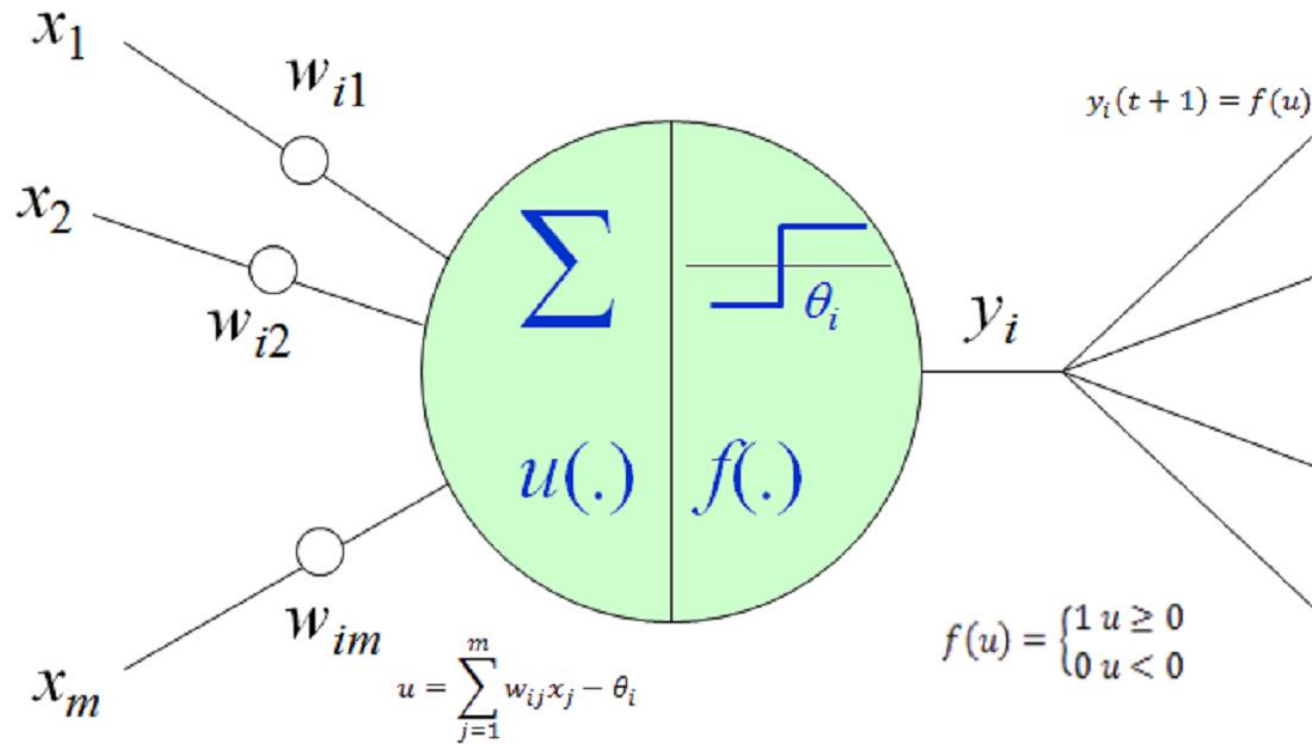
Αλγόριθμοι μάθησης



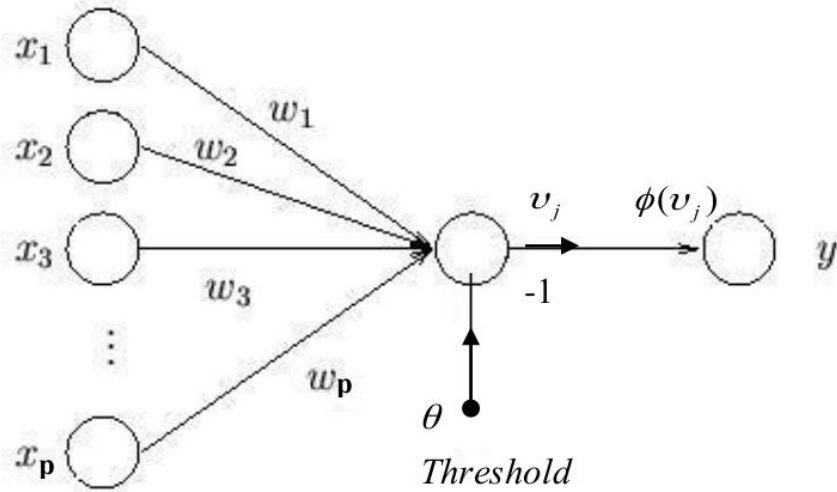
Ένα τέτοιο δίκτυο φαίνεται στο σχήμα που ακολουθεί με x_1 έως x_n να είναι η είσοδος, Y η έξοδος και b το κατώφλι (threshold).



Αλγόριθμοι μάθησης



Αλγόριθμοι μάθησης



Η έξοδος του γραμμικού συνδυαστή υπολογίζεται

$$v = \sum_{i=1}^p w_i x_i - \theta$$

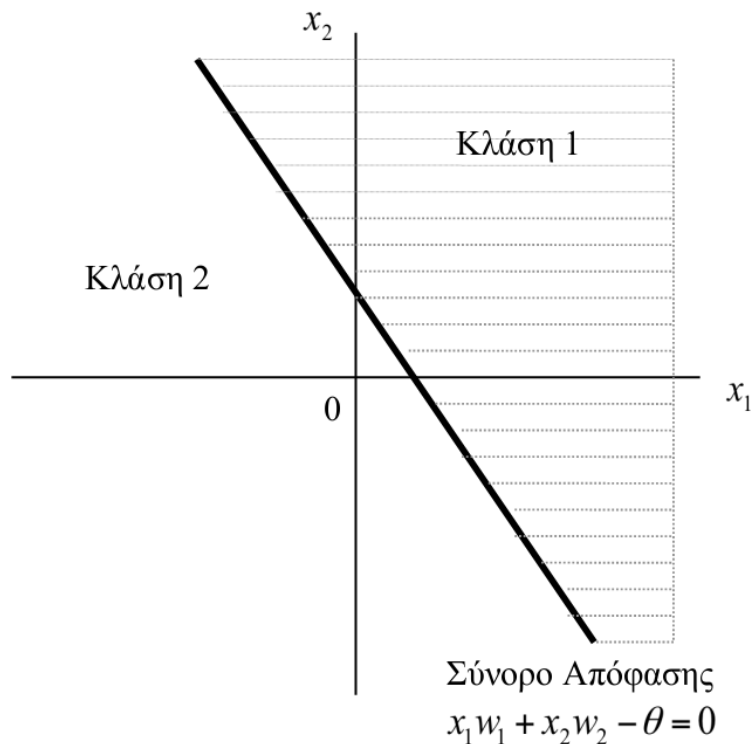
Σκοπός του Perceptron είναι να ταξινομήσει ένα σύνολο εισόδων (προτύπων) x_1, x_2, \dots, x_n σε μία από τις κλάσεις l_1 και l_2 . Ο κανόνας απόφασης για την ταξινόμηση είναι ο εξής:

ανάθεσε το σημείο που αναπαριστούν x_1, x_2, \dots, x_n στην κλάση l_1 , αν $y = +1$ και στην κλάση l_2 αν $y = -1$. Οι περιοχές απόφασης διαχωρίζονται από το υπερεπίπεδο που ορίζεται από τη σχέση:

$$v = \sum_{i=1}^p w_i x_i - \theta = 0 \quad \Leftrightarrow \quad x_1 w_1 + x_2 w_2 - \theta = 0$$

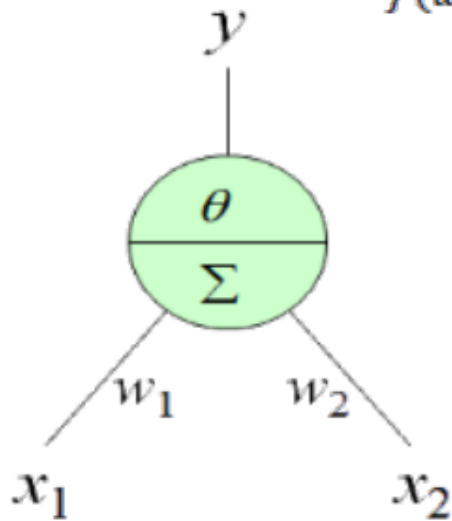
Αλγόριθμοι μάθησης

η γραμμική διαχωρισσιμότητα για ένα δισδιάστατο πρόβλημα ταξινόμησης, με δύο κλάσεις.



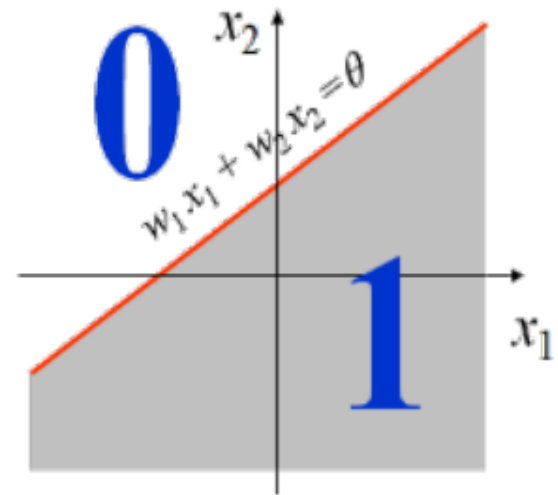
Από το παραπάνω σχήμα φαίνεται το αποτέλεσμα της εφαρμογής του κατωφλίου, το οποίο μετατοπίζει το όριο απόφασης από την αρχή των αξόνων. Τα συναπτικά βάρη του Perceptron, μπορούν να προσαρμοσθούν επαναληπτικά. Για την προσαρμογή του διανύσματος βαρών w , χρησιμοποιούμε έναν κανόνα διόρθωσης λάθους, που είναι γνωστός σαν κανόνας σύγκλισης του Perceptron.

Παράδειγμα



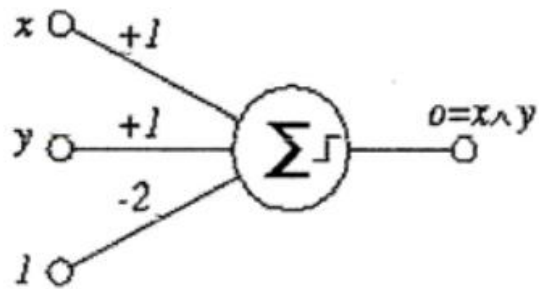
$$f(u) = w_1x_1 + w_2x_2 - \theta$$

$$y = \begin{cases} 1 & f(u) \geq 0 \\ 0 & f(u) < 0 \end{cases}$$

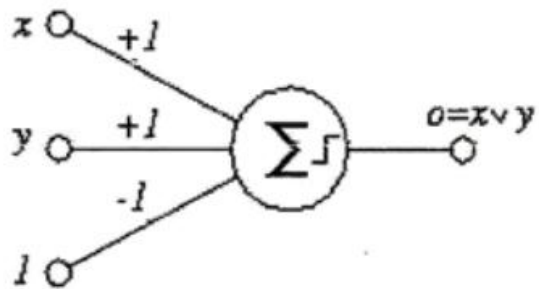


Παράδειγμα

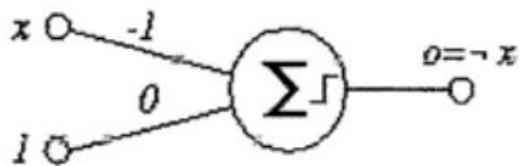
a. Λογικό ΚΑΙ



b. Λογικό Ή



c. Αναστροφή



Η βεβαρημένη συνάρτηση για το λογικό AND είναι:

$$\text{net} = 1 * x + 1 * y + (-2) * 1 = x + y - 2$$

x	y	$\text{net} = x + y - 2$	$f(\text{net}) = 0$
1	1	0	1
1	0	-1	0
0	1	-1	0
0	0	-2	0

Αλγόριθμοι μάθησης

ο αλγόριθμος σύγκλισης για ένα Perceptron n εισόδων με ένα υπολογιστικό νευρώνα, ακολουθεί τα παρακάτω βήματα:

$$x(n) = [1, x_1(n), x_2(n), \dots, x_n(n)]^T$$

$$w(n) = [b, w_1(n), w_2(n), \dots, w_n(n)]^T$$

Βήμα 1: Πρώτα υφίσταται αρχικοποίηση του διανύσματος των βαρών στο 0, δηλαδή $w(0)=0, n=1$

Βήμα 2: Στην συνέχεια, το μοντέλο Perceptron ενεργοποιείται στο χρονικό βήμα n , εφαρμόζοντας το διάνυσμα προτύπων εισόδων.

Βήμα 3: Με βάση τις τιμές των διανυσμάτων εισόδων και βαρών, υπολογίζεται η πραγματική έξοδος του δικτύου με χρήση των τύπων:

$$u(n) = w(n) \cdot x(n)$$

$$y(n) = \text{sgn}(u(n)) = \text{sgn}(w(n) \cdot x(n))$$

, όπου sgn είναι η συνάρτηση προσήμου $\text{sgn}(u) = +1$ αν $u > 0$ ή -1 αν $u \leq 0$

Αλγόριθμοι μάθησης

Βήμα 4: Μετά τον υπολογισμό της πραγματικής απόκρισης του δικτύου, για το διάνυσμα εισόδων $x(n)$, ο νευρώνας του Perceptron, προχωράει στην μεταβολή των βαρών,

με βάση τον κανόνα: $w(n+1) = w(n) + \eta \cdot [d(n) - y(n)] \cdot x(n)$, όπου “ η ” είναι η παράμετρος μάθησης (βήμα ή ρυθμός εκπαίδευσης) που ρυθμίζει το μέγεθος της διόρθωσης που θα επιβληθεί στα βάρη και παίρνει τιμές στο διάστημα $[0, 1]$. $d(n) = +1$ αν $x(n)$ ανήκει στην κλάση C_1 ή $d(n) = -1$ αν $x(n)$, ανήκει στην κλάση C_2 .

Βήμα 5: Τέλος, υλοποιείται αύξηση του χρονικού βήματος (επανάληψης) n κατά ένα και το δίκτυο επιστρέφει στο Βήμα 2 για να συνεχίσει την εκπαίδευσή του.

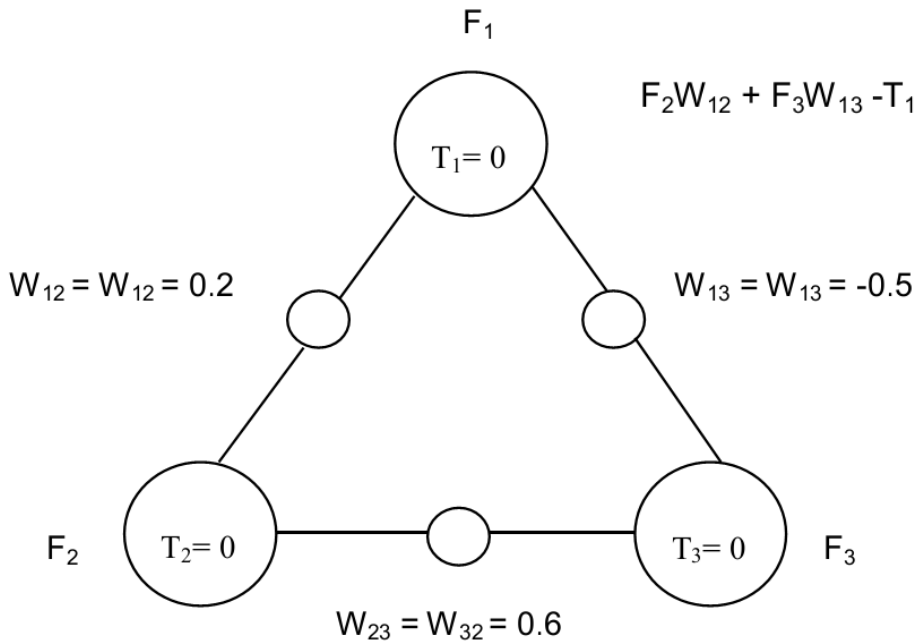
Παράδειγμα

Θα εξετάσουμε την παρούσα κατάσταση να είναι 010 αρχικά, δηλαδή:

$$F_1 = 0 \quad F_2 = 1 \quad F_3 = 0$$

Αν ο νευρώνας 1 είναι ο επόμενος για να προσπαθήσει να εκπέμψει, τότε μπορούμε να προβλέψουμε εάν θα εκπέμψει ή όχι εξετάζοντας τον κανόνα ενεργοποίησης για τον νευρώνα 1. Το άθροισμα των βαρών εισόδων συμπεριλαμβανομένου του κατώτατου ορίου - κατωφλίου (γνωστό ως ενεργοποίηση) για τον νευρώνα 1 φαίνεται να είναι :

$$F_2W_{12} + F_3W_{13} - T_1 = (1 * 0.2) + (0 * (-0.5)) + 0.1 = 0.3$$



Παράδειγμα

Η τιμή αυτή είναι μεγαλύτερη από το μηδέν, και ως εκ τούτου ο νευρώνας 1 θα εκπέμψει ($F_1 = 1$). Έτσι οι αλλαγές στο δίκτυο δηλώνουν ότι αν η κατάσταση είναι η 110 ο F_1 είναι ο επόμενος νευρώνας για να προσπαθήσει να εκπέμψει (πιθανότητα του 1 / 3).

Ωστόσο, πρέπει να εξετάσουμε επίσης τις περιπτώσεις κατά τις οποίες ο νευρώνας 2 ή ο νευρώνας 3 είναι ο επόμενος που θα προσπαθήσει να εκπέμψει, από την τωρινή κατάσταση 010. Αν ο νευρώνας 2 είναι ο επόμενος, η ενεργοποίησή του είναι :

$$F_1W_{21} + F_3W_{23} - T_2 = (0 * 0.2) + (0 * 0.6) - 0 = 0$$

Παράδειγμα

Η ενεργοποίηση είναι ίση με το μηδέν και συνεπώς, ($F_2 = 0$). Έτσι, η επόμενη κατάσταση από την 010 είναι η 000, αν ο F_2 επιχειρήσει να εκπέμψει και πάλι με πιθανότητα $1 / 3$.

Αν ο νευρώνας 3 είναι ο επόμενος για να προσπαθήσει να εκπέμψει, η ενεργοποίησή του είναι:

$$F_1W_{31} + F_2W_{32} - T_3 = (0 * (-0.5)) + (1 * 0.6) - 0 = 0.6$$

Η ενεργοποίηση είναι μεγαλύτερη από το μηδέν και συνεπώς, ($F_3=1$). Έτσι η επόμενη κατάσταση από το 010 είναι η 011 αν ο F_3 επιχειρήσει να εκπέμψει και πάλι με πιθανότητα $1 / 3$.