ε.δε.μ²
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Επισκόπηση Νευρωνικών Δικτύων
Κανόνας του Hebb
Ρύθμιση Παραμέτρων με Επιβλεπόμενη Μάθηση
Back Propagation Algorithm

καθ. Βασίλης Μάγκλαρης
maglaris@netmode.ntua.gr
www.netmode.ntua.gr

Αίθουσα 002, Νέα Κτίρια ΣΗΜΜΥ

Τρίτη 7/3/2023

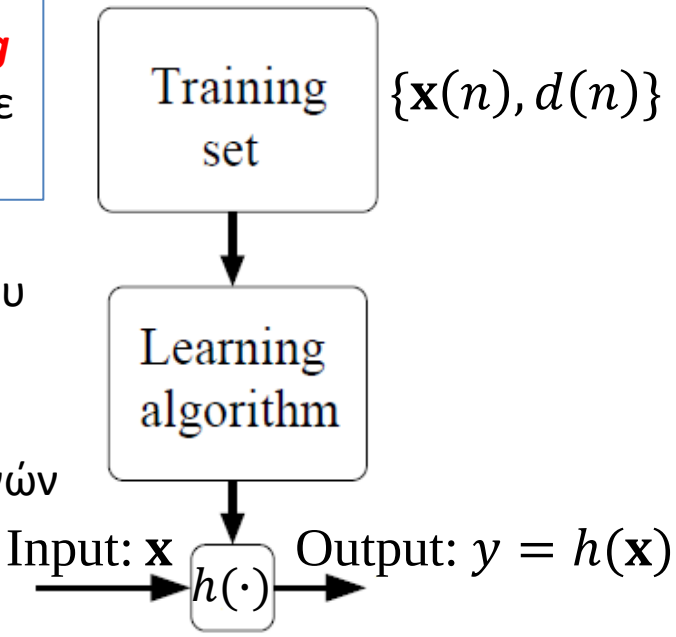
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning (*επανάληψη*)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- Στόχος του συστήματος είναι η αντιστοίχηση ενός δειγματικού στοιχείου εισόδου (***input sample point***, ***example***, ***instance***) $\mathbf{x} = [x_1\ x_2\ \dots x_m]^{\mathrm{T}}$ σε τιμές εξόδου y που εκτιμούν επιθυμητές τιμές d (***labels***, ***targets***) π.χ. πρόβλεψη ή ταξινόμηση. Τα στοιχεία x_i είναι αριθμητικές τιμές που κωδικοποιούν m ειδοποιά χαρακτηριστικά (***features***) του δειγματικού στοιχείου $\mathbf{x}$

Ζητείται ο προσδιορισμός της συνάρτησης εισόδου - εξόδου $y = h(\mathbf{x}) \cong d$ που προκύπτει από δείγμα μάθησης (***Training Set***) N ***labeled*** ζευγών $\{\mathbf{x}(n), d(n)\}$, $n = 1,2, \dots, N$ γνωστών σε εξωτερικό εκπαιδευτή (***supervisor***)



- Η μορφή και οι παράμετροι της $h(\cdot)$ προσδιορίζονται με αλγόριθμο μάθησης που συγκλίνει σε προσέγγιση του στόχου της υπόθεσης για τα N στοιχεία του δείγματος μάθησης

$$d(n) \cong y(n) = h(\mathbf{x}(n))$$

- Αν ο στόχος ικανοποιείται με μικρό αριθμό διακριτών επιλογών (κλάσεων) της y πρόκειται για πρόβλημα Ταξινόμησης, ***Classification*** (για δύο κλάσεις έχουμε δυαδική ταξινόμηση)
- Αν η έξοδος y λαμβάνει συνεχείς τιμές, το πρόβλημα αναφέρεται σαν Παλινδρόμηση, ***Regression***

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μοντέλα Μηχανικής Μάθησης (*επανάληψη*)

Διακριτικά Μοντέλα (**Discriminative Models**):

Μέθοδοι ταξινόμησης (*classification*) ή εκτίμησης (παλινδρόμηση, *regression*) δειγματικών στοιχείων (*data elements*) μέσω υπό συνθήκη πιθανότητας (*conditional density*) εξόδου (*label*) βάσει χαρακτηριστικών (*features*) του, όπως αυτές προσεγγίστηκαν σε στοιχεία δείγματος μάθησης (*training sample*) για γενίκευση σε *test datasets* (*generalization*)

Ενδεικτικές Εφαρμογές:

- *Ταξινόμηση δειγματικών στοιχείων* με βάση συνάρτηση χαρακτηριστικών τους
- *Αναγνώριση προτύπων* με βάση κύρια χαρακτηριστικά τους (*pattern recognition*)
- *Εκτίμηση εξόδου* συμβατή με διαθέσιμα ζεύγη εισόδου - στόχου (*regression*)

Παραγωγικά Μοντέλα (**Generative Models**):

Μέθοδοι εκτίμησης τρόπων παραγωγής (*generation*) δειγματικών στοιχείων, στατιστικά συμβατών με ιδιότητες του δείγματος μάθησης (*training sample*) μέσω συνδυασμένων πιθανοτήτων (*joint probabilities*) εξόδου (*output*) και χαρακτηριστικών (*features*) εισόδου, όπως υπολογίστηκαν στα στοιχεία μάθησης

Ενδεικτικές Εφαρμογές:

- *Δημιουργία προσομοιωμένων στοιχείων*: κειμένων (συμβατών με αποδεκτά μοντέλα Natural Language processing - NLP), εικόνων, κινούμενων σχεδίων, ιδεατών τοπίων…
- *Εμπλουτισμός Μηχανών Αναζήτησης* (*Google*, *MS Bing* + *OpenAI Chat Generative Pre-trained Transformer - ChatGPT*)
- *Επικράτηση αληθοφανών εναλλακτικών εκτιμήσεων* σε συνέργεια με εργαλεία θεωρίας παιγνίων (*Generative Adversarial Networks – GAN*)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός Παραμέτρων Linear Regression (1/2) (*επανάληψη*)

- Το διάνυσμα του δειγματικού στοιχείου εισόδου $\mathbf{x} = [x_0\ x_1\ \dots x_m]^{\mathrm{T}}$ ορίζεται με τιμές που κωδικοποιούν m χαρακτηριστικά (***features***) του: $x_1, x_2, \dots, x_m$ με $x_0 \triangleq 1$ (***intercept term***)
- Το σύστημα linear regression προσδιορίζει τις παραμέτρους $\mathbf{w} = [w_0\ w_1\ \dots w_m]^{\mathrm{T}}$ της συνάρτησης $y = h_{\boldsymbol{w}}(\mathbf{x}) = w_0 x_0 + w_1 x_1 + \dots + w_m\, x_m = \mathbf{w}^{\mathbf{T}}\mathbf{x}$ ώστε η y να έχει μικρές ***αποκλίσεις*** για το δείγμα μάθησης (***Training Set***) $\boldsymbol{\mathcal{D}} = \{(\mathbf{x}(1), d(1)), \dots, (\mathbf{x}(N), d(N))\}$
  - $\mathbf{x}(n)$: Διάνυσμα τιμών εισόδου (χαρακτηριστικών) στοιχείου μάθησης n (***regressors***)
  - $d(n)$: Τιμή εξόδου (*label*) στοιχείου μάθησης n (***regressand***)
  - $y(n) = h_{\boldsymbol{w}}(\mathbf{x}(n))$: Εκτίμηση εξόδου του συστήματος για διάνυσμα εισόδου $\mathbf{x}(n)$
  - $\varepsilon(n) = d(n) - y(n)$: Απόκλιση (***error***) εκτίμησης για το $\{\mathbf{x}(n), d(n)\}$, $n = 1,2, \dots, N$
  - Οι $\mathbf{x}(n), d(n), \varepsilon(n)$ μπορούν να θεωρηθούν δειγματικές τιμές τυχαίων μεταβλητών
- Κοινό κριτήριο σύγκλησης αφορά στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (***Least Mean Square***, ***LMS***) ως προς τις παραμέτρους $\mathbf{w}$ της $h_{\mathbf{w}}(\mathbf{x})$ ή αντίστοιχα της συνάρτησης κόστους:

$$J(\mathbf{w}) \triangleq \frac{1}{2}\sum_{n=1}^{N}[\varepsilon(n)]^2 = \frac{1}{2}\sum_{n=1}^{N}[d(n) - h_{\boldsymbol{w}}(\mathbf{x}(n))]^2$$

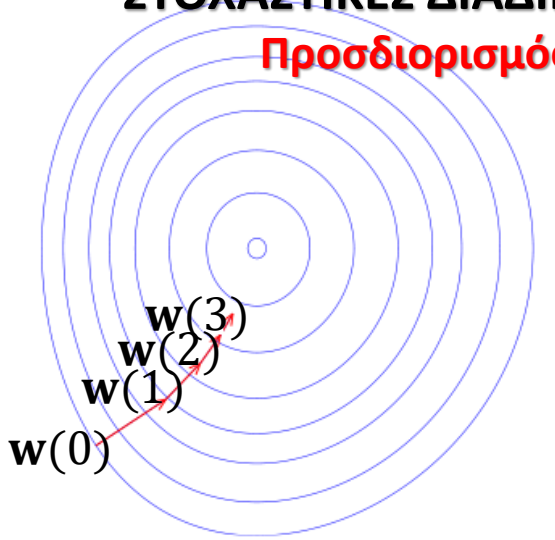
- Παράδειγμα **Linear Regression** μίας μεταβλητής εισόδου x:

$$\mathbf{x} = [1\ x]^{\mathrm{T}},\ \mathbf{w} = [w_0\ w_1]^{\mathrm{T}},\ \ \mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\,\mathbf{x} = w_0 + w_1 x$$

$$J(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}[d(n) - (w_0 + w_1 x(i))]^2$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός Παραμέτρων Linear Regression (2/2) (*επανάληψη*)



Απεικόνιση Σύγκλισης Gradient Descent:

Ελαχιστοποίηση της συνάρτησης $J(\mathbf{w})$, με παραμέτρους το διάνυσμα $\mathbf{w}$, μέσω διαδοχικής προσέγγισης στο βήμα $k \to k+1$ προς την κλίση (Gradient) $\nabla J(\mathbf{w})$ σταθμισμένο κατά την *hyperparameter* α:

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \alpha\, \nabla J(\mathbf{w}(k))$$

Αν υπάρχει σύγκλιση: $\mathbf{w} = \lim_{k\to\infty} \mathbf{w}(k)$

Σημείωση: Για linear regression υπάρχει πάντα σύγκλιση

Κανόνας Μάθησης LMS (Widrow-Hoff)

- **Batch Gradient Descent:** Προσδιορισμός του $\mathbf{w} = [w_0\ w_1 \dots w_m]^{\mathrm{T}}$ που ελαχιστοποιεί το σφάλμα $J(\mathbf{w})$ σε κάθε βήμα για ***όλο*** το δείγμα $\mathcal{D} = \{(\mathbf{x}(1), d(1)), \dots, (\mathbf{x}(N), d(N))\}$

$$w_j := w_j - \alpha \frac{\partial J(\mathbf{w})}{\partial w_j} = w_j + \alpha \sum_{n=1}^{N} [d(n) - h_{\mathbf{w}}(\mathbf{x}(n))]\, x_j(n), \quad j = 0,1,2,\dots,m \quad \forall i$$

- **Stochastic (Incremental) Gradient Descent, Stochastic Approximations**: Προσδιορισμός του $\mathbf{w}$ με τυχαία (στοχαστική) διαδοχική εισαγωγή ***στοιχείων*** $(\mathbf{x}(n), d(n)), n = 1,2,\dots,N$ του δείγματος μάθησης μέχρι να ικανοποιηθεί κριτήριο σύγκλησης

$$w_j := w_j + \alpha [d(n) - h_{\mathbf{w}}(\mathbf{x}(n))] x_j(n), \qquad j = 0,1,2,\dots,m \qquad n = 1,2,\dots,N$$

  - Η στοχαστική μέθοδος δίνει συνήθως ικανοποιητικά αποτελέσματα με μικρή επιβάρυνση υπολογιστικών πόρων και ***προτιμάται για μηχανική μάθηση***
  - Το βήμα α στις επαναλήψεις ορίζει τον ρυθμό της μάθησης (***learning rate***). Για σταθεροποίηση της σύγκλισης μπορεί να μεταβάλλεται στην πορεία των επαναλήψεων π.χ. μεγάλη τιμή στα πρώτα βήματα, μικρότερη όσο πλησιάζουμε στη σύγκλιση

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Polynomial Regression: Πολυωνυμική Προσέγγιση Χαρακτηριστικού (*επανάληψη*)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- Γραμμική προσέγγιση (**Linear Regression**) μίας μεταβλητής εισόδου $\mathbf{x} = [1\ x]^{\mathrm{T}}$:

$$\mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x$$

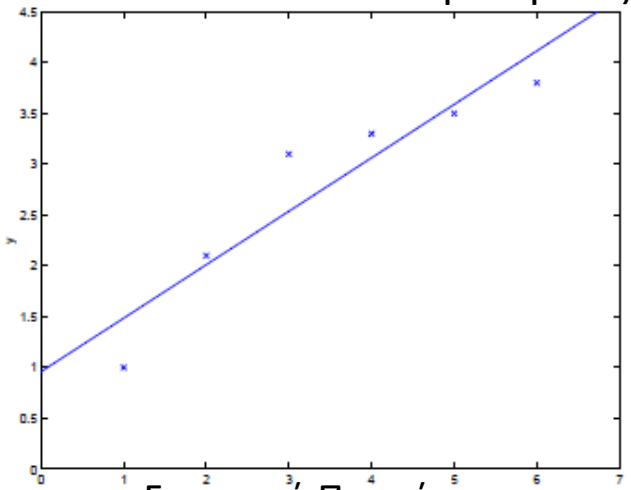
- Προσέγγιση με γραμμικό πολυώνυμο ***2ου βαθμού***:

$$\mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = w_0 + w_1 x + w_2 x^2$$

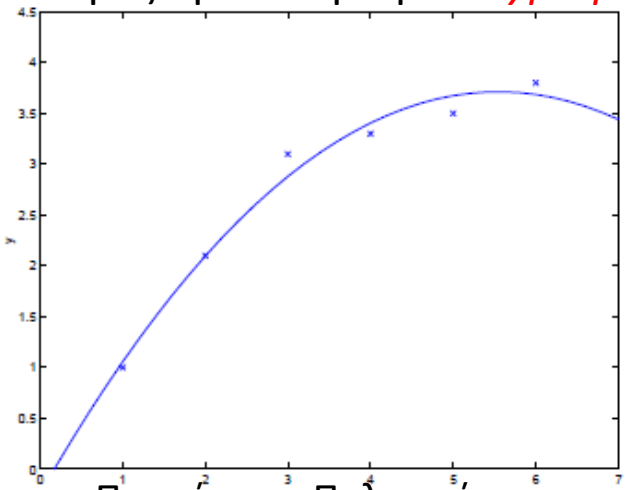
- Προσέγγιση με γραμμικό πολυώνυμο ***K* βαθμού** (*hyperparameter **K***):

$$\mathrm{y} = h_{\mathbf{w}}(\mathbf{x}) = \sum_{j=0}^{K} w_j x^j$$

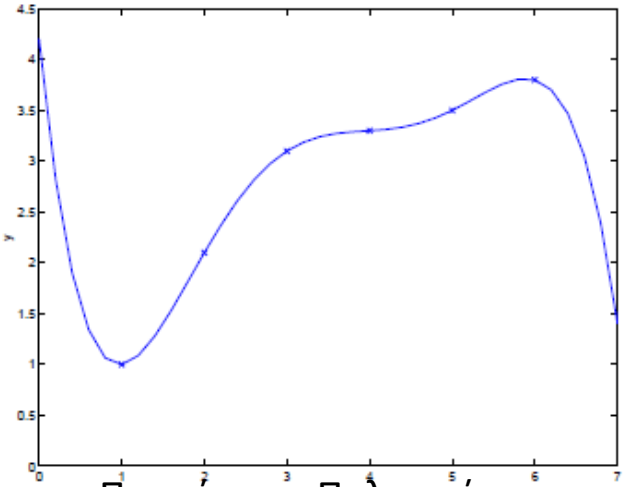
Εμπειρικές δοκιμές προσδιορισμού *hyperparameter **K***



Γραμμική Προσέγγιση
$K = 1$
(***Underfitting***)



Προσέγγιση Πολυωνύμου
2ου Βαθμού, $K = 2$
(***OK***)



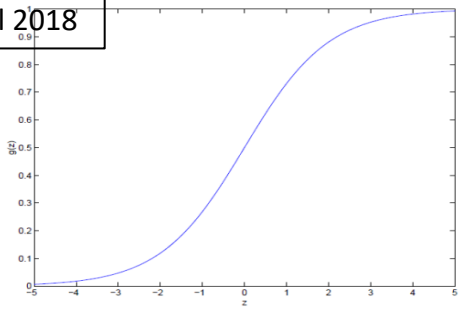
Προσέγγιση Πολυωνύμου
5ου Βαθμού, $K = 5$
(***Overfitting***)

Κίνδυνοι Υπεραπλούστευσης (***Underfitting***) & Υπερβολής (***Overfitting***)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ταξινόμηση – Classification (1/2) (επανάληψη)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018



Δειγματικά στοιχεία $\mathbf{x}$ με m διαστάσεις (χαρακτηριστικά, ***features***)
Δυαδικές Κλάσεις Εξόδου (***Classes***, ***Labels***) $y \in \{0,1\}$ ή $y \in \{-,+\}$
Training Set: $\{(\mathbf{x}(1), d(1)), \dots, (\mathbf{x}(N), d(N))\}$

- **Μοντέλο Logistic Regression**: $h_{\mathbf{w}}(\mathbf{x}) = g(\mathbf{w}^{\mathrm{T}}\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}^{\mathrm{T}}\mathbf{x}}}$

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} = \sum_{j=0}^{m} w_j\, x_j = w_0 + \sum_{j=1}^{m} w_j\, x_j$$

$$g(z) = \frac{1}{1+e^{-z}}$$
Logistic/Sigmoid Function

Οι πιθανότητες τυχαίας μεταβλητής εξόδου $y \in \{0,1\}$ υπό συνθήκη μεταβλητών εισόδου $\mathbf{x}$ και με συνάρτηση ***Logistic Regression*** $h_{\mathbf{w}}(\mathbf{x}) = \frac{1}{1+e^{-\mathbf{w}^{\mathrm{T}}\mathbf{x}}}$ ακολουθούν κατανομή ***Bernoulli*** και οδηγούν σε εκτίμηση της εξόδου y μετά τον προσδιορισμό των παραμέτρων $\mathbf{w}$:

$$P(y=1|\mathbf{x};\mathbf{w}) = h_{\mathbf{w}}(\mathbf{x}), \qquad P(y=0|\mathbf{x};\mathbf{w}) = 1 - h_{\mathbf{w}}(\mathbf{x})$$
$$\text{ή } p(y|\mathbf{x};\mathbf{w}) = \big(h_{\mathbf{w}}(\mathbf{x})\big)^{y}\big(1-h_{\mathbf{w}}(\mathbf{x})\big)^{1-y}$$

Κανόνας Εκτίμησης y :
$y = 1$ αν $h_{\mathbf{w}}(\mathbf{x}) > {}^{1}/_{2}$
$y = 0$ αν $h_{\mathbf{w}}(\mathbf{x}) < {}^{1}/_{2}$

Κριτήριο σύγκλησης λόγω μη γραμμικής $h_{\mathbf{w}}(\mathbf{x})$ προτιμάται της ελαχιστοποίησης του τετραγωνικού σφάλματος (***LMS***) η ***μεγιστοποίηση*** του λόγου πιθανοφάνειας (***Likelihood Ratio***) $L(\mathbf{w})$ των στοιχείων του συνόλου μάθησης $\{\mathbf{x}(n), d(n)\}, n = 1,2,\dots,N$. Θεωρούμε πως οι τιμές εξόδου $d(n)$ είναι ανεξάρτητες δυαδικές τυχαίες μεταβλητές για το δείγμα μάθησης $\mathbf{X} = [\mathbf{x}(1)\ \mathbf{x}(2)\ \dots\ \mathbf{x}(N)]$ και με παραμέτρους $\mathbf{w}$:

$$L(\mathbf{w}) \triangleq p\{(d(1), d(2), \dots, d(N))|(\mathbf{X};\mathbf{w})\} = \prod_{n=1}^{N} p\{(d(n)|(\mathbf{x}(n);\mathbf{w})\}$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ταξινόμηση – Classification (2/2) (*επανάληψη*)

Βασισμένο στο Andrew Ng, "CS229 Lecture Notes", Stanford University, Fall 2018

- **Μοντέλο Logistic Regression** (*συνέχεια*):

$$L(\mathbf{w}) = \prod_{n=1}^{N} p\{(d(n)|(\mathbf{x}(n);\mathbf{w})\} = \prod_{n=1}^{N} \left\{\left(h_{\mathbf{w}}(\mathbf{x}(n))\right)^{d(n)}\left(1-h_{\mathbf{w}}(\mathbf{x}(n))\right)^{1-d(n)}\right\}$$

Αντί της μεγιστοποίησης του $L(\mathbf{w})$ μεγιστοποιούμε τον λογάριθμο $l(\mathbf{w}) = \log L(\mathbf{w})$:

$$l(\mathbf{w}) = \sum_{n=1}^{N} \{d(n)\log h_{\mathbf{w}}(\mathbf{x}(n)) + (1-d(n))\log(1-h_{\mathbf{w}}(\mathbf{x}(n)))\}$$

Εφαρμόζουμε ***Gradient Ascent*** στο βήμα $k \to k+1$ με *hyperparameter* α θετική:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \alpha\, \nabla l(\mathbf{w}(k))$$

Για τον υπολογισμό της $\nabla l(\mathbf{w}(k))$ και την εφαρμογή του στοχαστικού προσεγγιστικού κανόνα (***Stochastic Gradient Ascent***) προσδιορισμού των παραμέτρων w_j με διαδοχική εφαρμογή στα στοιχεία $n = 1,2,\dots,N$ του ***Training Set*** υπολογίζουμε την μερική παράγωγο $\frac{\partial}{\partial w_j} l(\mathbf{w}) = \cdots = \left[d(n) - h_{\mathbf{w}}\left(\mathbf{x}(n)\right)\right]x_j(n)$ ⇨

$$w_j := w_j + \alpha\left[d(n) - h_{\mathbf{w}}\left(\mathbf{x}(n)\right)\right]x_j(n), \quad j = 1,2,\dots,m$$

(ίδιας μορφής ***Επαναληπτικός Κανόνας Μάθησης*** με τον κανόνα ***LMS***)

- **Μοντέλο Perceptron**: $h_{\mathbf{w}}(\mathbf{x}) = g\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}\right)$

$$g(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$ **Threshold Function**

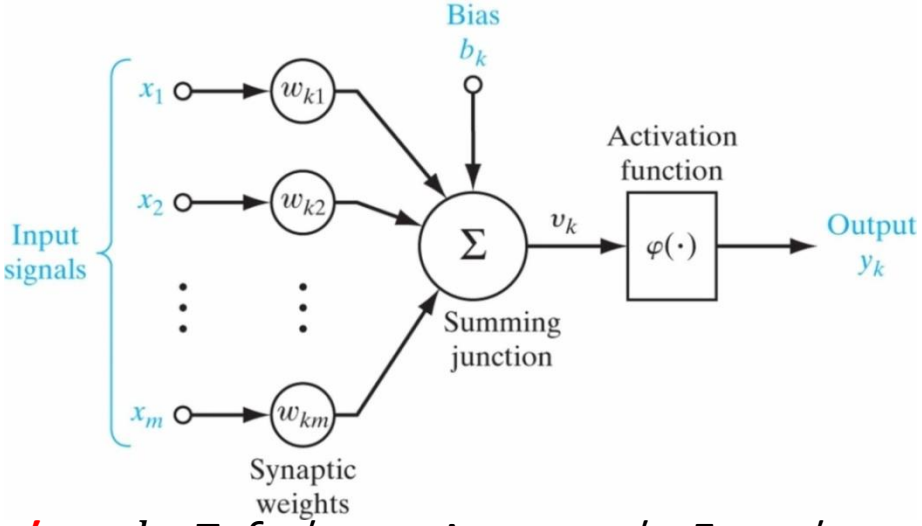
Προκύπτει παρόμοιος ***Επαναληπτικός Κανόνας Μάθησης*** παραμέτρων w_j

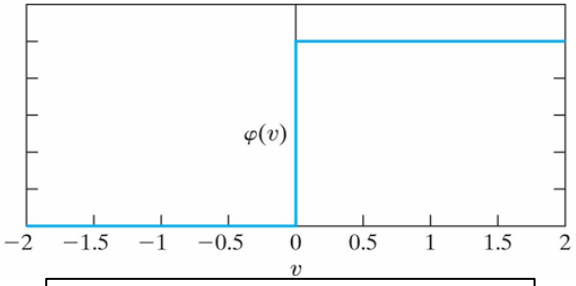
$$w_j := w_j + \alpha\left[d(n) - h_{\mathbf{w}}\left(\mathbf{x}(n)\right)\right]x_j(n), \quad j = 1,2,\dots,m \quad \text{για} \quad n = 1,2,\dots,N$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μη Γραμμικό Μοντέλο Δυαδικού Τεχνητού Νευρώνα k: *Rosemblatt's* Perceptron

McCulloch & ***Pitts*** (1943): Νευρωνικά Δίκτυα σαν μηχανές Μηχανικής Μάθησης
Hebb (1949): Αρχές αυτοοργάνωσης μαθησιακής διαδικασίας
Rosemblatt (1958): Επιβλεπόμενη μάθηση, Perceptron
Rumelhart (1985): Back Propagation Algorithm





Συνάρτηση Πρόσημου
(***Signum Function*** sgn)
$\varphi(v) \in \{-1, +1\}$

Νευρώνας k **:** Ταξινόμηση Δειγματικών Στοιχείων $\mathbf{x} = [x_1\ x_2\ \dots x_m]^{\mathrm{T}}$ σε 2 Κλάσεις $\{-1, +1\}$

Input Signals (Features of input Sample Points): $x_j \triangleq \pm 1,\ j = 1,2, \dots, m$

Synaptic Weights: Παράμετροι συνάψεων $\mathbf{w}_k = [w_{k0}\ w_{k1} \dots w_{km}]^{\mathrm{T}}$

Bias: $b_k \triangleq w_{k0}$, Intercept term $x_0 \triangleq +1$

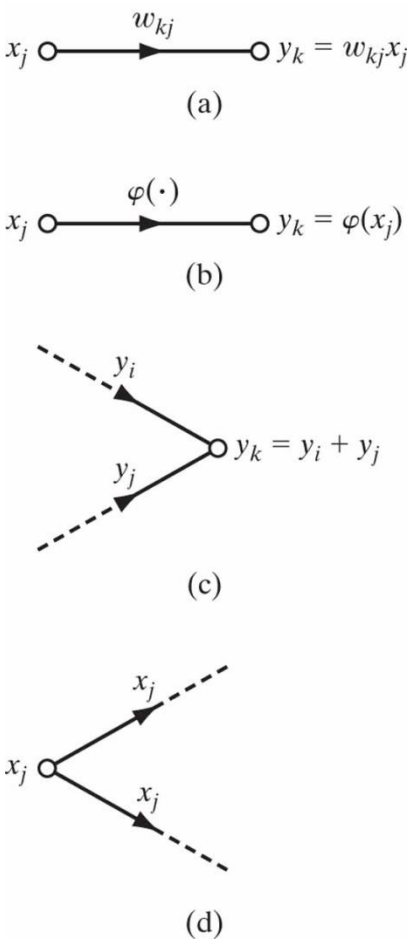
Induced Local Field - Activation Potential: $v_k = \sum_{j=0}^{m} w_{kj} x_j$

Δυαδική Έξοδος: $y_k = \varphi(v_k) = \mathrm{sgn}(v_k) \in \{-1(Inactive), +1(Active)\}$

Επιβλεπόμενη Μάθηση: Ρύθμιση w_{kj} από N labeled στοιχεία training dataset $\{\mathbf{x}(n), d(n)\}$ για ελαχιστοποίηση αποκλίσεων, π.χ. Mean Square Error (***MSE***): $\min_{\mathbf{w}_k}\{\frac{1}{N}\sum_{n=1}^{N}[d(n) - y_k(n)]^2\}$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

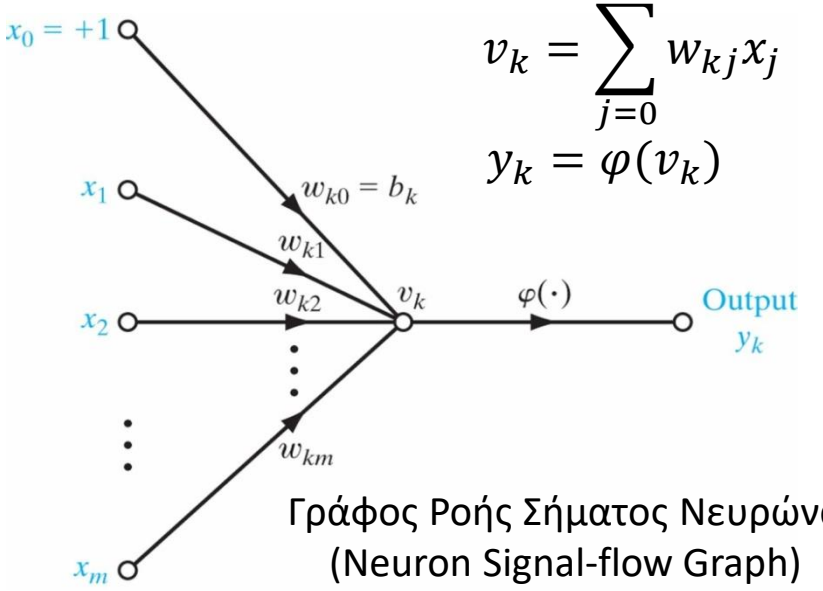
Νευρωνικά Δίκτυα σαν Κατευθυνόμενοι Γράφοι



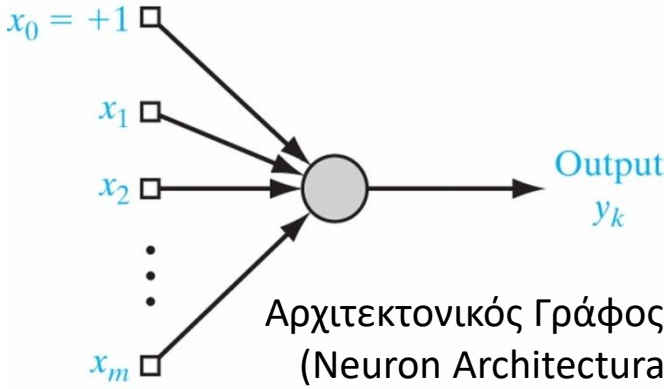
Κανόνες Γράφων Ροής Σήματος
(Signal-flow Graph Rules)

$$v_k = \sum_{j=0}^{m} w_{kj}x_j$$

$$y_k = \varphi(v_k)$$



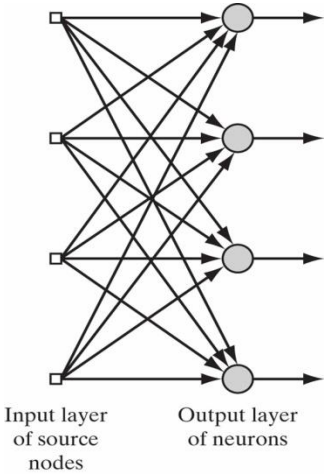
Γράφος Ροής Σήματος Νευρώνα
(Neuron Signal-flow Graph)



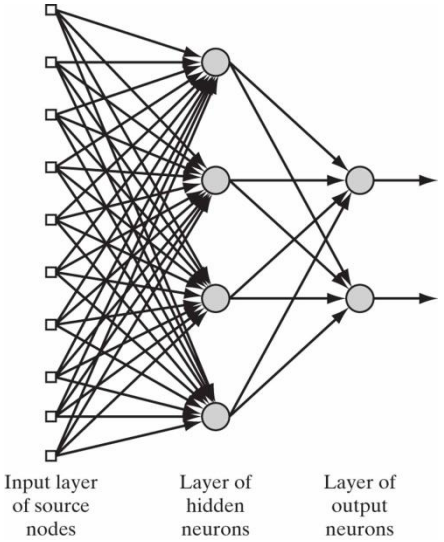
Αρχιτεκτονικός Γράφος Νευρώνα
(Neuron Architectural Graph)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

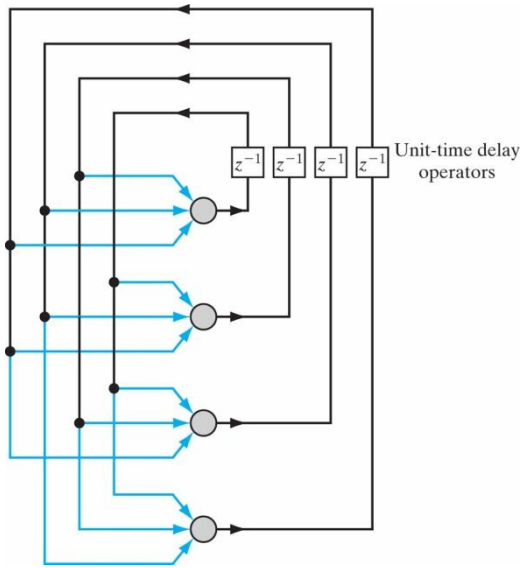
Μοντέλα Νευρωνικών Δικτύων



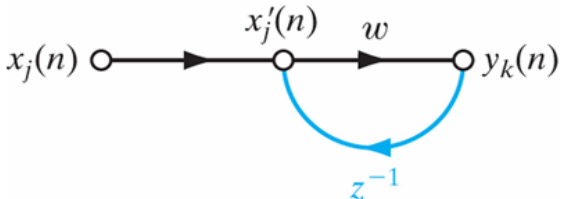
Μονοστρωματικό Δίκτυο Πρόσθιας Τροφοδότησης
(Single-Layer Feedforward Network)



Δίκτυο Πρόσθιας Τροφοδότησης με Κρυφούς Νευρώνες
(Multilayer Feedforward Network with Hidden Neurons)



Αναδρομικό Δίκτυο Hopfield
(Recurrent Network with no Hidden Neurons)



Signal-flow Graph Φίλτρου IIR 1ης Τάξης με Ανάδραση
(Infinite Impulse Response - IIR Filter with Feedback)

$x_j(n)$: Σήμα Εισόδου στην διακριτή στιγμή n

$y_k(n)$: Σήμα Εξόδου στην διακριτή στιγμή n

$x_j'(n)$: Εσωτερικό Σήμα στην διακριτή στιγμή n

$$y_k(n) = \sum_{l=0}^{\infty} w^{l+1} x_j(n-l)$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Φάσεις Σχεδίασης Νευρωνικών Δικτύων (ΝΔ)

- Φάση Μάθησης (***Learning***) με χρήση γνωστών δεδομένων δείγματος μάθησης (*training sample points*) από το περιβάλλον:
  - Προσδιορισμός *synaptic weights* w_{kj} και *biases* b_k σε συγκεκριμένο μοντέλο ΝΔ
  - Επαναληπτικοί αλγόριθμοι μάθησης (π.χ. με κριτήριο *Mean Square Error* - *MSE*)
- Φάση Επικύρωσης (***Validation***): Επαλήθευση διακριτικής ικανότητας μοντέλων ΝΔ μέσω πρόσθετων γνωστών δεδομένων επικύρωσης (*validation sample points*):
  - Επιλογή *hyperparameters* (αριθμός νευρώνων, στρωμάτων, κριτήρια σύγκλισης)
  - Αποφυγή *overfitting*
- Φάση Ελέγχου (***Testing***) επίδοσης (ακρίβεια) σε νέα δεδομένα (*testing sample points*):
  - Αξιολόγηση γενίκευσης (*generalization*) της διακριτικής ικανότητας επιλεγμένου ΝΔ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αναπαράσταση Γνώσης
(Knowledge Representation)

Διακριτική ικανότητα (προβλέψεις, ερμηνεία και επιθυμητή απόκριση) σε εξωτερικές εισόδους του **περιβάλλοντος** που θα λειτουργήσει ένα νευρωνικό δίκτυο με:

- Πρότερη Πληροφορία (**Prior Information**)

- Μετρήσεις – Παρατηρήσεις Παραδειγμάτων (**Observations**)
  - Συνήθως εμπεριέχουν θόρυβο (noisy sensor errors)
  - Παρέχουν υποδείγματα εισόδου για ***training*** (υπολογισμό παραμέτρων) νευρωνικών δικτύων

  - **Labeled** (αντιστοίχηση με επιθυμητά χαρακτηριστικών εξόδου με μεσολάβηση ***εκπαιδευτή***)
  - **Unlabeled** (χωρίς αντιστοίχηση χαρακτηριστικών εξόδου)

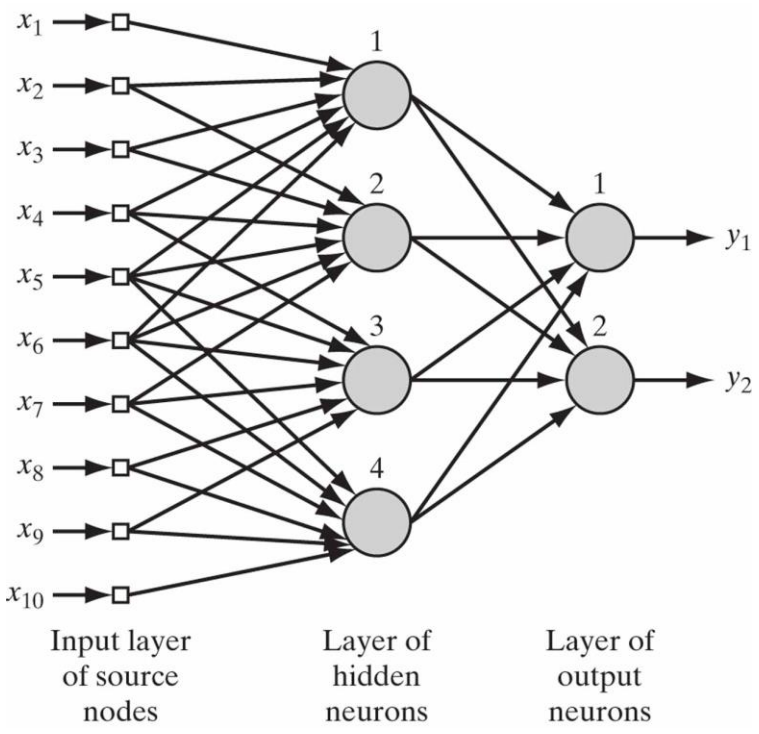
Κανόνας του *Hebb* (ομοιότητες με νευροφυσιολογικά μοντέλα, 1949)
Σε δίκτυα τεχνητών νευρώνων αναδεικνύονται τάσεις σταδιακής ενίσχυσης συνδέσεων μεταξύ ενεργών νευρώνων (synapses between active neurons), ανάλογα με αυτά που παρατηρούνται σε νευροφυσιολογικά συστήματα μάθησης.

Τα συναπτικά βάρη w_{ij} ανάμεσα σε ενεργά στοιχεία (νευρώνες) i, j τείνουν προς ενίσχυση ενώ οι άλλες συνάψεις τείνουν προς μηδενισμό. Ο κανόνας αυτός κωδικοποιεί ένα οδηγό αυτορρύθμισης περίπλοκων συστημάτων μηχανικής μάθησης π.χ. μη επιβλεπόμενη μάθηση σε Self-Organizing Maps, Boltzmann Machines κλπ.

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ενσωμάτωση Πρότερης Πληροφορίας στη Σχεδίαση Νευρωνικών Δικτύων (Building Prior Information into Neural Network Design)

Γενικές επιλογές αρχιτεκτονικής με οδηγό Πρότερη Πληροφορία συγκεκριμένης εμπειρίας



Επιλογές Απλοποίησης Αρχιτεκτονικής Νευρωνικού Δικτύου:

1. Αρχιτεκτονική με ορισμό πεδίου υποδοχής (*receptive field*) ίσου αριθμού εισόδων ανά κρυφό νευρώνα
2. Διαμοιρασμός βαρών (*weight sharing*): Στο παράδειγμα με 10 κόμβους εισόδου και ένα στρώμα 4 κρυφών νευρώνων, ορίζονται κοινά βάρη $w_i, i = 1, \dots, 6$ για συνάψεις 6 εναλλακτικών επιλογών εισόδων από τις x_i, $i = 1, \dots, 10$ προς τους κρυφούς νευρώνες 1,2,3,4

$$v_j = \sum_{i=1}^{6} w_i x_{i+j-1}, \qquad j = 1,2,3,4$$

Συνελικτικά αθροίσματα (*Convolutional Neural Network*)

Induced Local Fields – Activation Potential Κρυφών Νευρώνων

$$v_1 = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$
$$v_2 = w_1x_2 + w_2x_3 + w_3x_4 + w_4x_5 + w_5x_6 + w_6x_7$$
$$v_3 = w_1x_4 + w_2x_5 + w_3x_6 + w_4x_7 + w_5x_8 + w_6x_9$$
$$v_4 = w_1x_5 + w_2x_6 + w_3x_7 + w_4x_8 + w_5x_9 + w_6x_{10}$$

Ο υπολογισμός των βαρών w_i απαιτεί διαδικασία Μάθησης που θα προσδιορίσει τελική λύση

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Σχεδίαση Νευρωνικών Δικτύων

Επιλογές Αρχιτεκτονικής

- Στρώματα (layers – input, output, hidden)
- Πρόσθια τροφοδότηση (feedforward)
- Ανάδραση (feedback, recurrent)

Προσδιορισμός Παραμέτρων, Μάθηση

- Επιβλεπόμενη Μάθηση με Εκπαιδευτή - **Supervised Learning**
  - Με χρήση **Labeled Training Sample Points** (ζεύγη εισόδου – εξόδου) για ρύθμιση της αρχιτεκτονικής με κριτήριο την ελαχιστοποίηση κόστους πρόβλεψης

- Μάθηση χωρίς Εκπαιδευτή - **Unlabeled Training Sample Points**
  - Μη Επιβλεπόμενη Μάθηση - **Unsupervised Learning**
  - Ενισχυτική Μάθηση - **Reinforcement Learning**

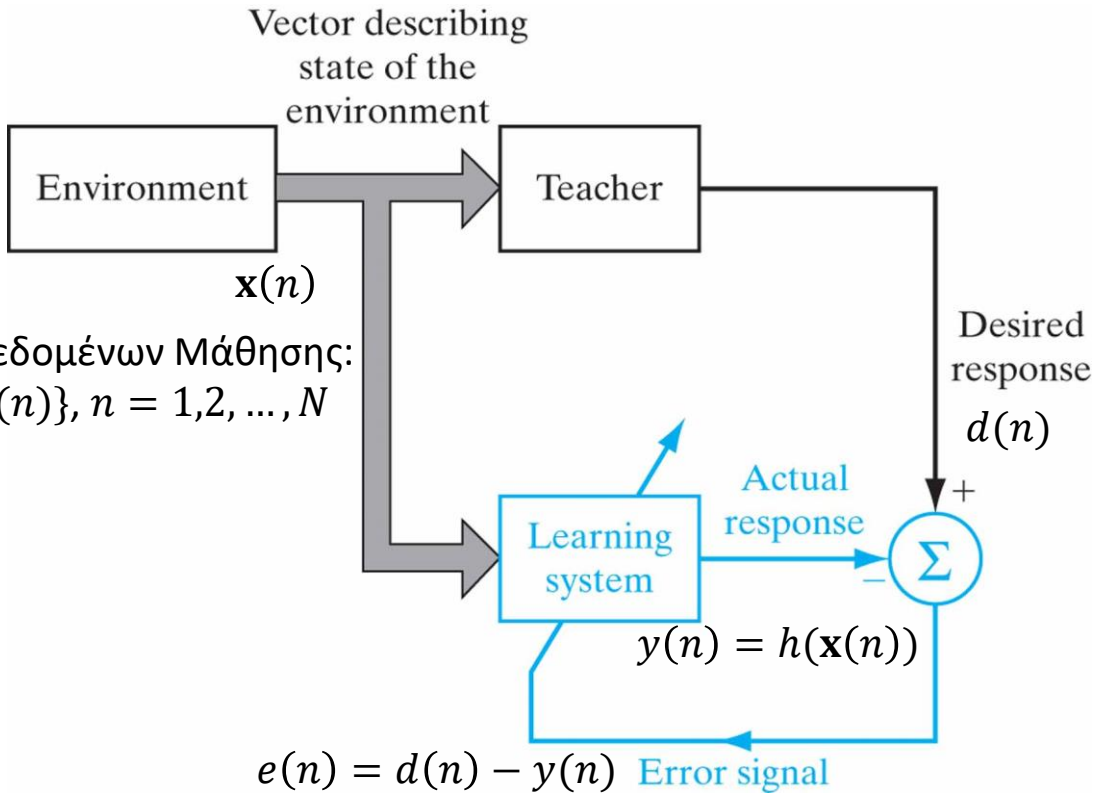
- Κανονικοποίηση των Δεδομένων Μάθησης - **Normalization of Training Datasets**
  - Feature Scaling, **Min-max Normalization** πριν την εφαρμογή βελτιστοποίησης

- Ανανέωση των Παραμέτρων με βάση το Δείγμα Μάθησης προς αποδοτική Σύγκλιση
  - Είτε **On-line** μετά από είσοδο κάθε δειγματικού στοιχείου μάθησης (π.χ. **Stochastic Gradient Descent**) είτε μετά από είσοδο του συνόλου (**Batch**) ή υποσυνόλων (**Mini-Batches**) του δείγματος μάθησης (π.χ. **Batch Gradient Descent**)
  - Δυνατότητα επαναλήψεων κατά εποχές (**Epochs**) για το σύνολο των δεδομένων μάθησης
    https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μάθηση με Εκπαιδευτή

Supervised Learning



Νευρωνικό Δίκτυο (Learning System)

- Εκτίμηση βαθμωτής (scalar) συνάρτησης $y = h(\mathbf{x})$ της κατάστασης δειγματικών στοιχειών $\mathbf{x}$ του περιβάλλοντος (***Environment State Vector***) π.χ. ταξινόμηση σε δύο κλάσεις με βάση δυαδική συνάρτησή συνιστωσών (χαρακτηριστικών) του $\mathbf{x}$

Ζεύγη Δεδομένων Μάθησης: $\{\mathbf{x}(n), d(n)\}, n = 1,2, \dots, N$

Φάση Μάθησης

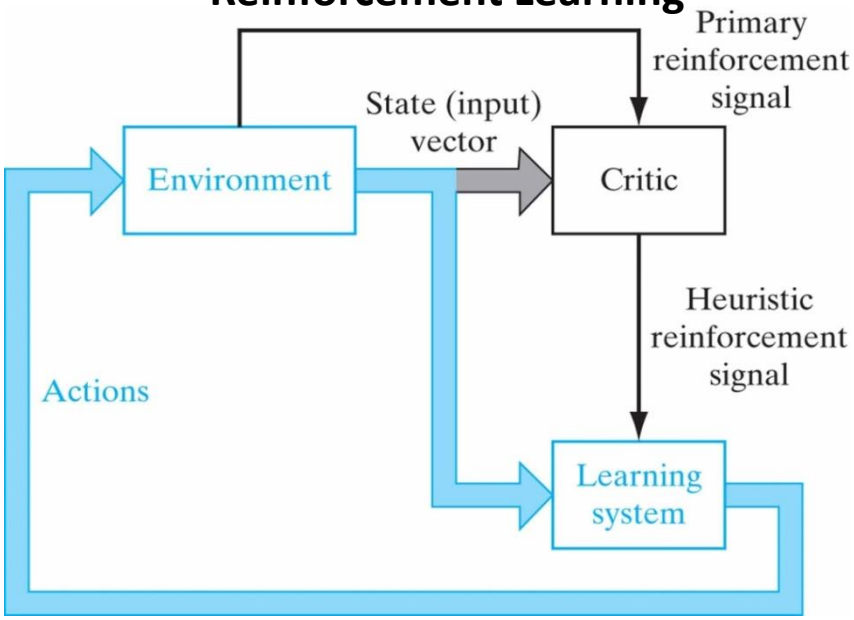
- Δείγμα μάθησης με στοιχεία την κατάσταση του περιβάλλοντος $\mathbf{x}(n)$ και την αντίστοιχη ζητούμενη έξοδο (desired response, label) $d(n)$, γνωστή στον Εκπαιδευτή (***Teacher***)
- Εκτίμηση εξόδου $y(n) = h(\mathbf{x}(n))$ από το Νευρωνικό Δίκτυο (***Actual Response***), υπολογισμός απόκλισης $e(n) = d(n) - y(n)$ από ***Desired Response***, διόρθωση παραμέτρων συστήματος σε επαναλήψεις για μείωση του σφάλματος $e(n)$
- Άμεση σχέση με μεθόδους ***στατιστικής*** εκτίμησης και ***βελτιστοποίησης*** με επαναληπτικούς αλγορίθμους

Το Learning System ρυθμίζει τις παραμέτρους του $h(\cdot)$ ***επαναληπτικά με οδηγό την απόκλιση (σφάλμα) της εξόδου του*** $y(n) = h(\mathbf{x}(n))$ ***από την ζητούμενη απόκριση*** $d(n)$***, γνωστή στον Εκπαιδευτή, π.χ. σύμφωνα με μέθοδο Steepest Descent προς την κλίση (gradient) συναρτήσεως σφάλματος όπως αυτό εξελίσσεται σαν ακολουθία (Στοχαστική Διαδικασία) κατά την είσοδο των Δεδομένων Μάθησης*** $n = 1,2, \dots, N$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μάθηση χωρίς Εκπαιδευτή – Ενισχυτική Μάθηση

Reinforcement Learning

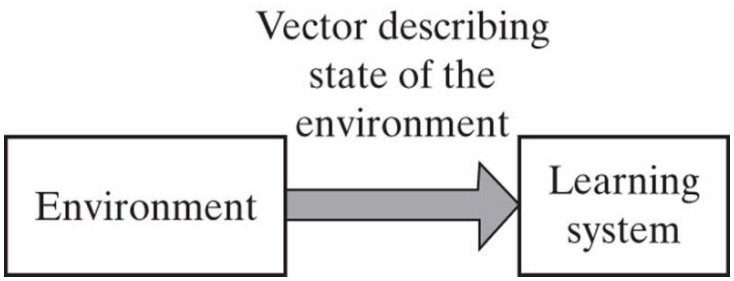


- Το Σύστημα (***Learning System***) μαθαίνει από το περιβάλλον χωρίς ***labeled training data*** και εκπαιδευτή. Αλλά παίρνει υπόψη του πρόσθετα ενισχυτικά σήματα από το περιβάλλον μέσω εξωτερικού κριτή (***Critic***, ***Agent***) και επιδρά στην εξέλιξη του περιβάλλοντος μέσω ενεργειών (***Actions***)
- Το περιβάλλον υπολογίζει ένα βαθμωτό σήμα ενίσχυσης της απόδοσης του συστήματος (***Scalar Primary Reinforcement Signal***) το οποίο διαβάζει στον κριτή (***Critic***, ***Agent***) μαζί με τη κατάσταση εισόδου (***State Vector***) σε κάθε επανάληψη κατά την εξέλιξη της κατάστασής του
- Ο εξωτερικός κριτής υπολογίζει ***πολιτικές*** με εκτίμηση κόστους/επιβράβευσης κατά την αναμενόμενη εξέλιξη της κατάστασης του περιβάλλοντος. Υπολογίζει ένα απλοποιημένο βαθμωτό σήμα ενίσχυσης (***Heuristic Reinforcement Signal***) για την υλοποίηση της πολιτικής και το διαβιβάζει στο Σύστημα
- Το Σύστημα ρυθμίζει τις παραμέτρους του ανάλογα με τη κατάσταση εισόδου και το απλοποιημένο σήμα ενίσχυσης. Μέσω βρόχου ανάδρασης (***Feedback***) προωθεί στο περιβάλλον ενέργειες (***Actions***) που επιδρούν στην εξέλιξη της κατάστασης για την επίτευξη μέσω-μακροπρόθεσμου στόχου (μεγιστοποίηση οφέλους ή ελαχιστοποίηση κόστους)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μάθηση χωρίς Εκπαιδευτή: Μη Επιβλεπόμενη Μάθηση

Unsupervised Learning

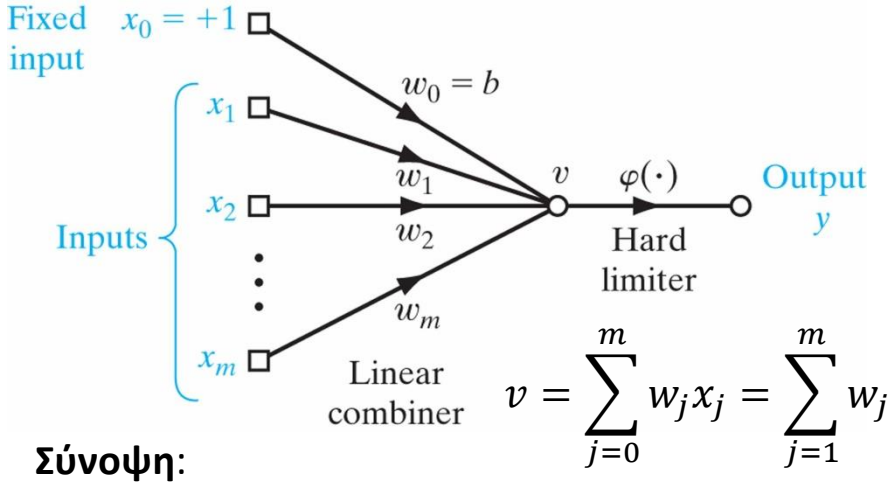


- Το σύστημα αυτορυθμίζεται ανακαλύπτοντας από μόνο του ***σημαντικές*** στατιστικές διακυμάνσεις του περιβάλλοντος. Υπολογίζει ενδιαφέρουσες στατιστικές δομές (***stochastic features***, ***patterns***) σε μεγάλο όγκο μη χαρακτηρισμένων δεδομένων (***unlabeled datasets***) από τα οποία προκύπτουν μοντέλα, μέθοδοι επεξεργασίας, αποθήκευσης και ταξινόμησης, π.χ. σε ομάδες (***clusters***)
- Το σύστημα μπορεί να ***δημιουργήσει*** δειγματικά στοιχεία (***generated sample elements***) συμβατά με τις στατιστικές ιδιότητες του περιβάλλοντος. Αποτέλεσμα: Η ταξινόμηση και συμπλήρωση ελλειμματικών περίπλοκων δεδομένων (π.χ. για επεξεργασία εικόνων και αναγνώριση προτύπων)
- Παράδειγμα ***unsupervised learning***: Νευρωνικό δίκτυο δυο επιπέδων, επίπεδο εισόδου και κρυφό επίπεδο από νευρώνες που ανταγωνίζονται για την αποθήκευση βασικών χαρακτηριστικών των δειγματικών στοιχείων εισόδου (***Competitive Learning***)
- Απλή υλοποίηση - επέκταση κανόνα του ***Hebb***: Κατά την μάθηση ενεργοποιείται ο νευρώνας με τη μεγίστη τιμή διέγερσης v_k (***winner-takes-all***)

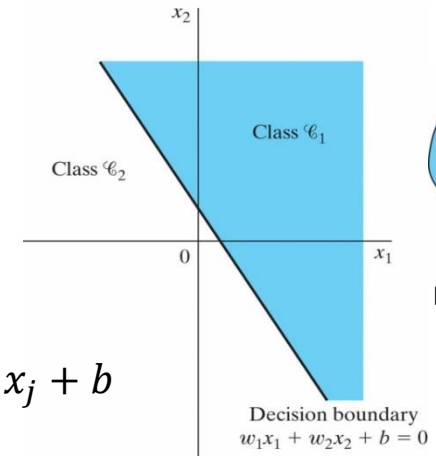
Πρόβλημα overfitting αν το σύστημα προσπαθεί να παρακολουθεί μη σημαντικές διακυμάνσεις

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

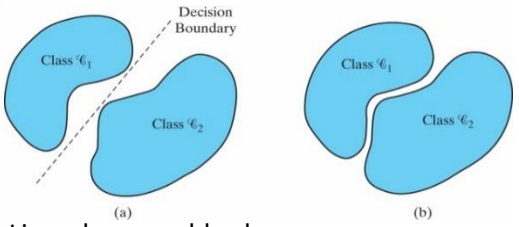
Rosenblatt's Perceptron



$$v = \sum_{j=0}^{m} w_j x_j = \sum_{j=1}^{m} w_j x_j + b$$



Non-linearly separable classes



Linearly separable classes

Σύνοψη:

1. Ένας νευρώνας με γραμμικό ***induced local field*** v και συνάρτηση ενεργοποίησης $\varphi(v)$ κατωφλίου (***Threshold Function***, ***Hard Limiter***) ή πρόσημου (***Signum Function***) για δυαδική ταξινόμηση στοιχείων $\mathbf{x} = [x_0\ x_1\ \dots x_m]^{\mathrm{T}}$ σε δύο ***γραμμικά διαχωριζόμενες*** κλάσεις:

$$\mathcal{C}_1 \text{ αν } y = \varphi(v) = 1, \qquad \mathcal{C}_2 \text{ αν } y = \varphi(v) = 0 \text{ ή αν } y = \varphi(v) = -1$$

1. Τα βάρη $\mathbf{w} = [w_0\ w_1\ \dots w_m]^{\mathrm{T}}$ ρυθμίζονται on-line (stochastic iterative method) με την εφαρμογή ***Error-correction Algorithm*** σε δειγματικά στοιχεία μάθησης $\{\mathbf{x}(n), d(n)\}, n = 1,2, \dots, N$ σε περιβάλλον ***supervised learning*** προς ελαχιστοποίηση σφαλμάτων $[d(n) - y(n)]$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta[d(n) - y(n)]\mathbf{x}(n)$$

Η *hyperparameter* $\eta, 0 < \eta \leq 1$ (*learning-rate parameter*) αν είναι *μικρή* οδηγεί την επαναληπτική διαδικασία μάθησης σε σύγκλιση. Αν είναι *μεγάλη* μπορεί να επιταχύνει τη σύγκλιση π.χ. σε περιβάλλοντα με μεγάλες αποκλίσεις των δεδομένων $\mathbf{x}(n)$, αλλά μπορεί να οδηγήσει σε αστάθειες λόγω ταλαντώσεων περί την βέλτιστη τιμή

Σε περιβάλλον δειγματικών στοιχείων $\mathbf{x}$ κατανομής Gauss, η ταξινόμησή τους σε δύο κλάσεις $\mathcal{C}_1, \mathcal{C}_2$ μέσω **Bayes Classifier** (ελαχιστοποίηση ρίσκου σφάλματος με βάση a-priori πιθανότητες p_1, p_2) ταυτίζεται με το **Rosenblatt Perceptron**

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

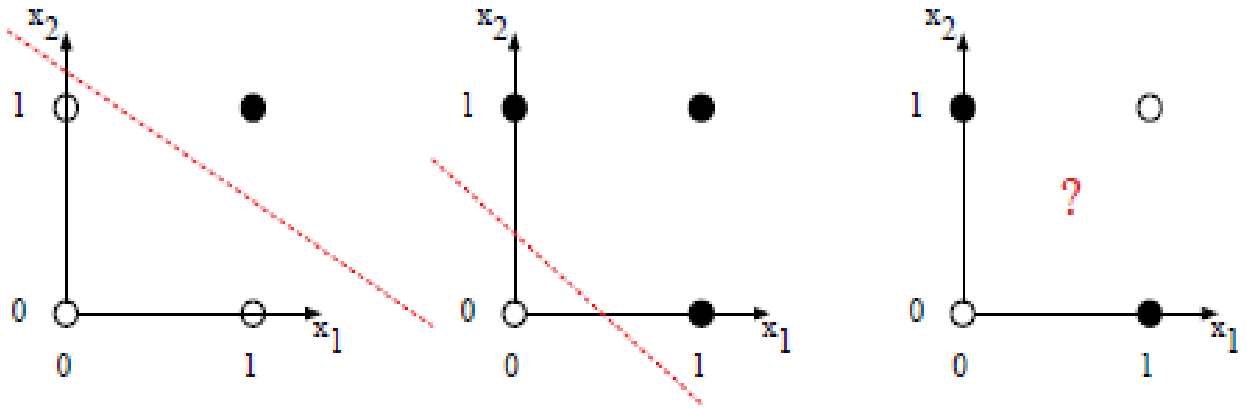
Ταξινόμηση με Perceptron Δυαδικών Εισόδων: Έξοδος Συναρτήσεις AND, OR (*XOR* ?)

AND			OR			XOR		
x1	x2	y	x1	x2	y	x1	x2	y
0	0	0	0	0	0	0	0	0
0	1	0	0	1	1	0	1	1
1	0	0	1	0	1	1	0	1
1	1	1	1	1	1	1	1	0

Συνάρτηση Διέγερσης Κατωφλίου (*threshold*):

$$y = h(v) = \begin{cases} 0, v \leq 0 \\ 1, v > 0 \end{cases}$$

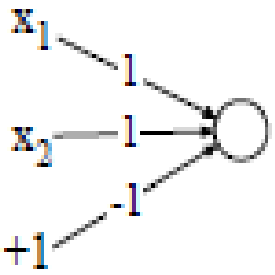
$$y = \begin{cases} 0, & w_1 \cdot x_1 + w_2 \cdot x_2 + b \leq 0 \\ 1. & w_1 \cdot x_1 + w_2 \cdot x_2 + b > 0 \end{cases}$$

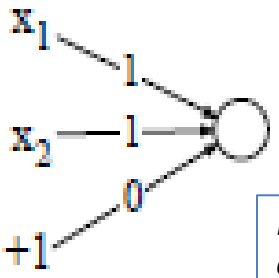


a) x_1 AND x_2 b) x_1 OR x_2 c) x_1 XOR x_2

Γραμμικές Καμπύλες Διαχωρισμού

Μη Γραμμική Καμπύλη Διαχωρισμού





Αδυναμία υλοποίησης με Perceptron ενός επιπέδου

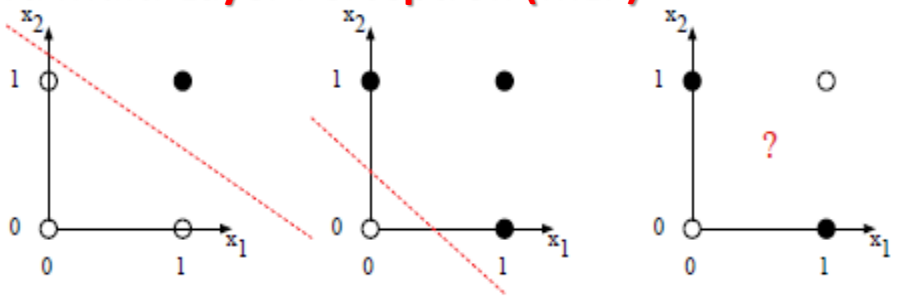
Βασισμένο στο Daniel Jurafsky, James H. Martin, ***"Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition"****, Third Edition draft, 2018*

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ταξινόμηση με Perceptron Δυαδικών Εισόδων: Έξοδος Συνάρτηση XOR

Υλοποίηση με Feedforward Multi-Layer Perceptron (MLP)

AND			OR			XOR		
x1	x2	y	x1	x2	y	x1	x2	y
0	0	0	0	0	0	0	0	0
0	1	0	0	1	1	0	1	1
1	0	0	1	0	1	1	0	1
1	1	1	1	1	1	1	1	0

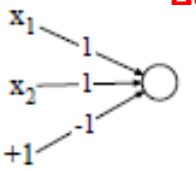


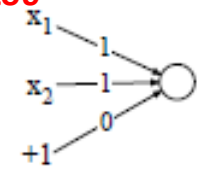
Γραμμικές Καμπύλες Διαχωρισμού

Μη Γραμμική Καμπύλη Διαχωρισμού

Συνάρτηση Διέγερσης: $y = h(v) = \begin{cases} 0, v \leq 0 \\ 1, v > 0 \end{cases}$

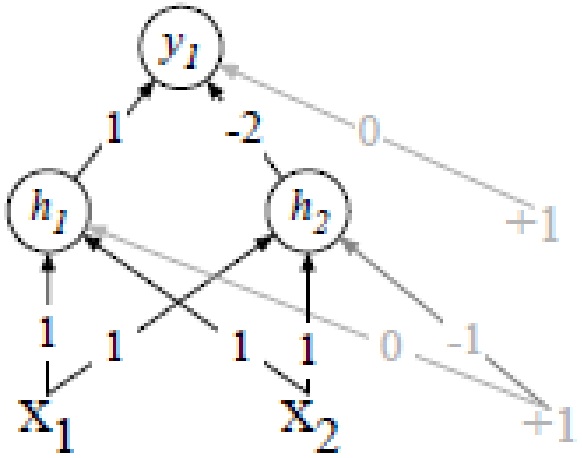
$$y = \begin{cases} 0, & w_1 \cdot x_1 + w_2 \cdot x_2 + b \leq 0 \\ 1, & w_1 \cdot x_1 + w_2 \cdot x_2 + b > 0 \end{cases}$$





Αδυναμία υλοποίησης με Perceptron ενός επιπέδου

Υλοποίηση XOR: Κρυφό Επίπεδο Νευρώνων $h_1, h_2 \rightarrow$ **Multi-Layer Perceptron, MLP**



Έξοδος Νευρώνα h_1: $a_1 = h(w_{11} \cdot x_1 + w_{21} \cdot x_2 + b_1)$

Έξοδος Νευρώνα h_2: $a_2 = h(w_{12} \cdot x_1 + w_{22} \cdot x_2 + b_2)$

Έξοδος Νευρώνα y_1: $y = h(w_{13} \cdot a_1 + w_{23} \cdot a_2 + b_3)$

$w_{11} = w_{12} = w_{21} = w_{22} = 1, b_1 = 0, b_2 = -1$

$w_{13} = 1, w_{23} = -2, b_3 = 0$

x_1: 0 0 1 1
x_2: 0 1 0 1
a_1: 0 1 1 0
a_2: 0 0 1 0
y: 0 1 1 0

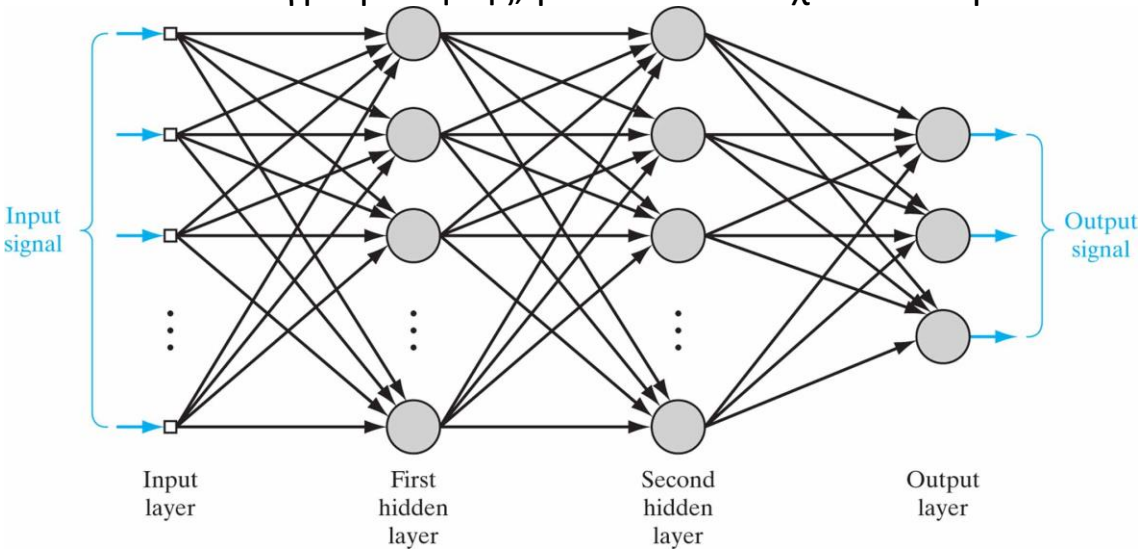
Βασισμένο στο Daniel Jurafsky, James H. Martin, **"Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition"**, Third Edition draft, 2018

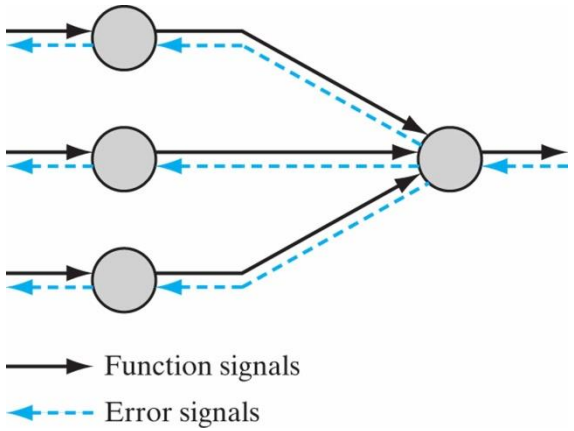
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Multilayer Perceptrons , Back-Propagation Algorithm (1/3)

Χαρακτηριστικά:

1. Διαφορίσιμες μη γραμμικές συναρτήσεις ενεργοποίησης $\varphi_j(v_j)$ του νευρώνα j (π.χ. ***Logistic Function***)
2. Το δίκτυο περιλαμβάνει κρυφά επίπεδα (***hidden layers***) νευρώνων με μεγάλη συνδεσιμότητα και συναπτικά βάρη w_{ji} (από $i \to j$). Οι κρυφοί νευρώνες ενισχύουν ειδοποιά χαρακτηριστικά (features) του δείγματος εισόδου μέσω διαδικασίας επιβλεπόμενης μάθησης
3. Τα βάρη ρυθμίζονται με την εφαρμογή ***Back-Propagation Algorithm***, συνήθως ***on-line*** (***stochastic iterative method***) διαδοχικά για κάθε ***labeled*** παράδειγμα μάθησης $\{\mathbf{x}(n), d_j(n)\}, n = 1,2,\dots,N$ σε περιβάλλον ***supervised learning*** με δύο φάσεις:
   i. **Forward Phase**: Η είσοδος $\mathbf{x}(n)$ του παραδείγματος μάθησης n διαπερνά το δίκτυο μέσω ***Function Signals*** με βάρη $w_{ji}(n)$ όπως έχουν προσδιοριστεί μέχρι την εφαρμογή του παρόντος παραδείγματος και προκύπτει η έξοδος $y_j(n)$ του νευρώνα j
   ii. **Backward Phase**: Οι αποκλίσεις $e_j(n) = d_j(n) - y_j(n)$ διαπερνούν το δίκτυο στην ανάστροφη πορεία σαν ***Error Signals*** και διορθώνουν τα συναπτικά βάρη
4. Η τελική σύγκλιση ολοκληρώνεται σε πολλαπλές ***εποχές*** με επαναλήψεις των δύο φάσεων για το ***συνολικό*** δείγμα μάθησης, με τα ίδια στοιχεία αλλά μετά από τυχαία ανακατανομή της σειράς τους





ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Multilayer Perceptrons , Back-Propagation Algorithm (2/3)

Σύγκληση κατά Least Mean Square – LMS ανά Εποχή :

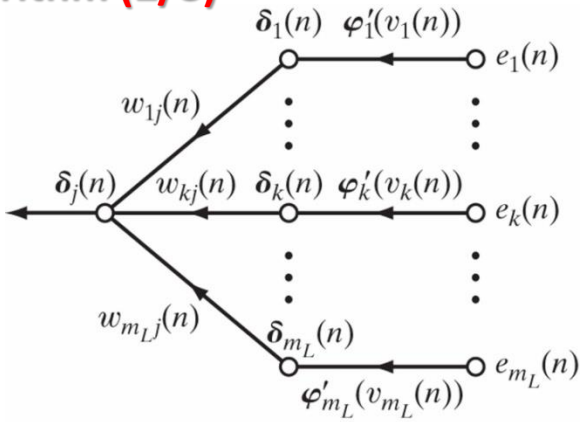
Για το παράδειγμα μάθησης $\{\mathbf{x}(n), d_j(n)\}$ το σφάλμα στους νευρώνες εξόδου j είναι $e_j(n) = d_j(n) - y_j(n)$. Το μέσο τετραγωνικό σφάλμα ή ο *Μέσος Όρος Στιγμιαίων Ενεργειών Αποκλίσεων* είναι $\boldsymbol{\mathcal{E}}(n) = \frac{1}{2}\sum_j e_j^2(n)$ και για όλο το δείγμα μάθησης σε μια *Εποχή* προκύπτει η μέση τιμή $\boldsymbol{\mathcal{E}}_{AVG}(N), n = 1,2,\dots,N$:

$$\boldsymbol{\mathcal{E}}_{AVG}(N) = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{\mathcal{E}}(n) = \frac{1}{N}\sum_{n=1}^{N}\left\{\frac{1}{2}\sum_j e_j^2(n)\right\}$$

Οι επαναληπτικές διορθώσεις $\Delta w_{ji}(n)$ στα βάρη $w_{ji}(n)$ οδηγούν προς την ελαχιστοποίηση του $\boldsymbol{\mathcal{E}}(n)$ στη κατεύθυνση της *local gradient* $\delta_j(n) = \frac{\partial \boldsymbol{\mathcal{E}}(n)}{\partial v_j(n)}$

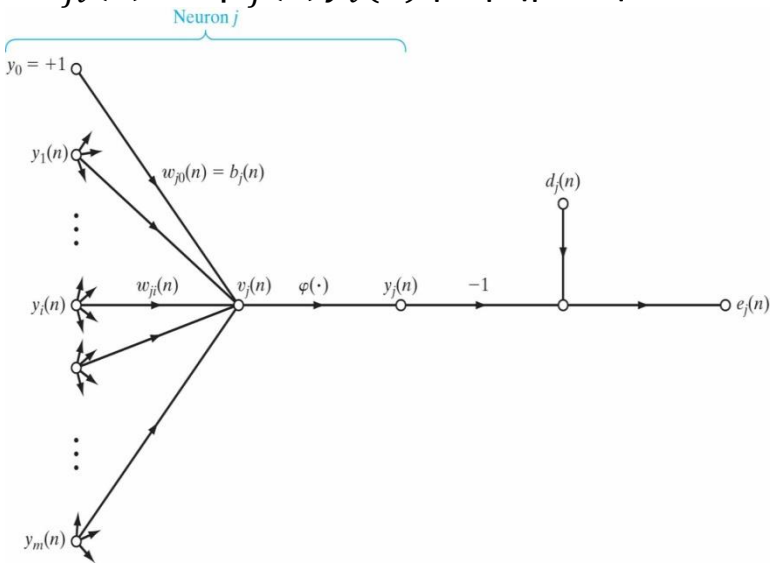
όπου $v_j(n) = \sum_{i=0}^{m} w_{ji}(n)y_i(n)$ (*Induced Local Field* του νευρώνα j):

$\Delta w_{ji}(n) = \eta\delta_j(n)y_i(n)$ με βήμα την *Learning Parameter* η

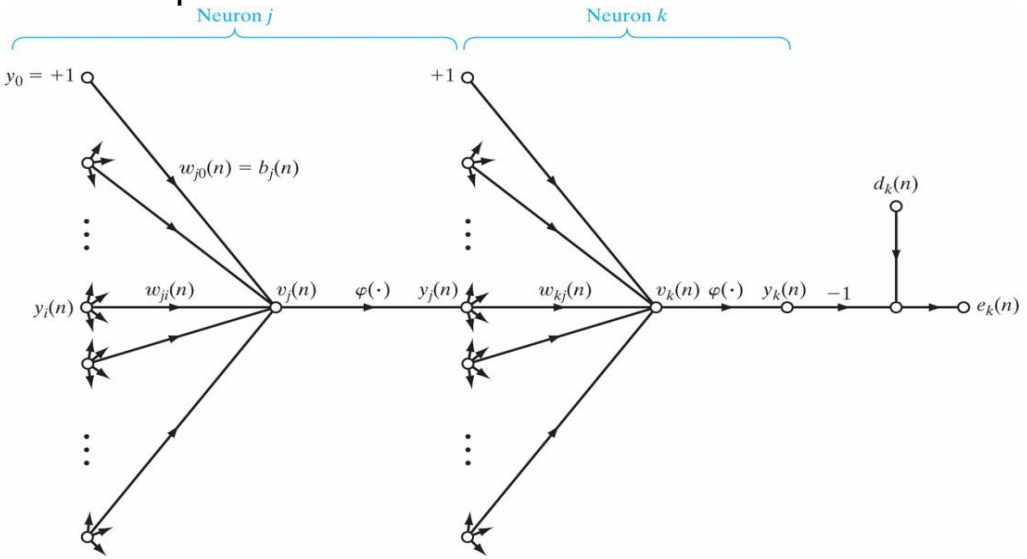


$$\varphi_j'\left(v_j(n)\right) = \frac{\partial y_j(n)}{\partial v_j(n)}$$

Ροή σήματος για ανάστροφη διάδοση σφαλμάτων



Ροή σήματος σε νευρώνα εξόδου j



Ροή σήματος νευρώνα εξόδου k συνδεόμενου σε κρυφό νευρώνα j

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Multilayer Perceptrons , Back-Propagation Algorithm (3/3)

Στην υλοποίηση ***on-line learning***, για κάθε παράδειγμα μάθησης n μετά από αρχικοποίηση των βαρών $w_{ji}^{(l)}(0)$ για κάθε layer $l = 1,2, \dots, L$ υλοποιούνται οι υπολογισμοί των φάσεων ***forward*** & ***backward***

Για το παράδειγμα μάθησης $\{\mathbf{x}(n), d_j(n)\}, n = 1,2, \dots, N$ με αποκρίσεις $d_j(n)$ του νευρώνα j:

1. Υπολογισμοί φάσης ***forward***:

   Για κάθε νευρώνα j και κάθε layer $l = 1,2, \dots, L$ το ***induced local field*** είναι $v_j^{(l)}(n) = \sum_i w_{ji}^{(l)}(n)\, y_j^{(l-1)}(n)$. Για $i = 0,\ y_0^{(l-1)} = +1, w_{j0}^{(l)} = b_j^{(l)}(n)$ οι τιμές bias του νευρώνα j

   Η ***έξοδος*** του κάθε νευρώνα είναι $y_j^{(l)}(n) = \varphi_j(v_j^{(l)}(n))$

   Για το πρώτο κρυφό επίπεδο $y_j^{(0)}(n) = x_j(n)$

   Αν το j ανήκει στο επίπεδο εξόδου $y_j^{(L)}(n) = o_j(n)$ όπου $o_j(n)$ είναι η τελική έξοδος j του Multilayer Perceptron. Το ***σήμα σφάλματος*** είναι $e_j(n) = d_j(n) - o_j(n)$

2. Υπολογισμοί φάσης ***backward***:

   Υπολογισμός τοπικών μερικών παραγώγων

$$\delta_j^{(l)}(n) = \begin{cases} e_j^{(L)}(n)\ \varphi_j'\left(v_j^{(L)}(n)\right) & \text{αν το } j \text{ ανήκει στο επίπεδο εξόδου} \\ \varphi_j'\left(v_j^{(L)}(n)\right) \sum_k \delta_K^{(l+1)}(n) w_{ki}^{(l+1)} & \text{αν το } j \text{ ανήκει σε κρυφό επίπεδο} \end{cases}$$

   Τα βάρη ρυθμίζονται με τον επαναληπτικό κανόνα με βήμα την ***Learning Hyperparameter*** η:

$$w_{ji}^{(l)}(n+1) = w_{ji}^{(l)}(n) + \alpha \times w_{ji}^{(l)}(n-1) + \eta \times \delta_j^{(l)}(n)(n) y_j^{(l-1)}(n)$$

   Η σταθερά $\alpha \geq 0$ (***momentum constant***) βοηθά στη σταθεροποίηση της σύγκλισης με συμμετοχή τιμών που προέκυψαν από προηγούμενα παραδείγματα μάθησης

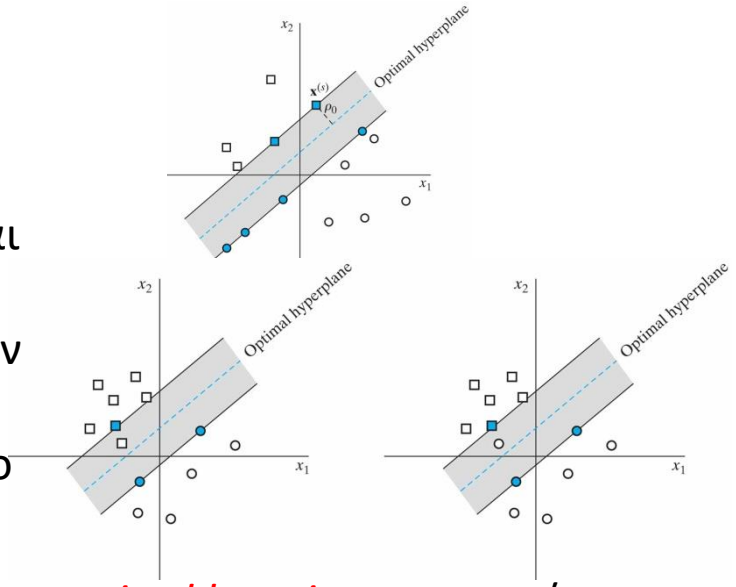
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αναφορά σε Ειδικές Μηχανές Μάθησης - Νευρωνικά Δίκτυα

Convolutional Neural Networks (CNN): Ειδική κατηγορία *Multilayer Perceptron* για αποδοτική κατηγοριοποίηση δυσδιάστατων δειγμάτων (π.χ. *pattern recognition* εικόνων) κυρίως με μάθηση μέσω δασκάλου (*supervised learning*). Η απλοποίηση προκύπτει με αποσύνδεση του δικτύου σε χαλαρά συνδεόμενα επίπεδα (*layers*), κοινά χαρακτηριστικά (*receptive fields*) και συνελικτικά αθροίσματα για συναρτήσεις διέγερσης (*convolutional induced local fields*)

K- Means Clustering: Οργανώνει σε K συστάδες (*clusters*) παρατηρήσεις $\mathbf{x}_i$ με βάση κοινά χαρακτηριστικά χωρίς δάσκαλο (*unsupervised learning*), π.χ. οργάνωση συστάδων με βάση την *Ευκλείδεια απόσταση* $\|\mathbf{x}_i - \mathbf{x}_j\|^2$

Support Vector Machines (SVM): Ταξινόμηση δεδομένων μέσω δύο περιοχών μεγίστης γραμμικής διάκρισης με *supervised learning*. Τα όρια των περιοχών αυτών ορίζονται από σημεία υποστήριξης (*support vectors*) όπως στο δυσδιάστατο σχήμα με τα μπλε τετράγωνα και κύκλους. Αν δεν υπάρχει όριο γραμμικής διάκρισης περιοχών (*non-separable patterns*), ζητείται μηχανή που προκύπτει από το δείγμα μάθησης με το ελάχιστο σφάλμα



Self-Organizing Maps (SOM): Βασίζεται σε *competitive unsupervised learning* που μειώνει τους ενεργοποιημένους νευρώνες όσο πλησιάζουν στο επίπεδο εξόδου. Οι νευρώνες τοποθετούνται στους κόμβους ενός δυσδιάστατου πλέγματος (*lattice*) και οργανώνονται σε τοπογραφικούς χάρτες μέσω παραδειγμάτων μάθησης που πυροδοτούν διαδρομές θετικής και αρνητικής ανάδρασης ώστε να προκύψει *τελικός νικητής* (*winner takes all*)