

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μη Επιβλεπόμενη Μάθηση - Unsupervised Learning

***K*-Means Clustering**

Ανάλυση Κυρίων Συνιστωσών - Principal Component Analysis (PCA)

Self-Organizing Maps (SOM)

Autoencoders

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Αίθουσα 02, Νέα Κτίρια ΣΗΜΜΥ

Τρίτη 15/3/2022

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

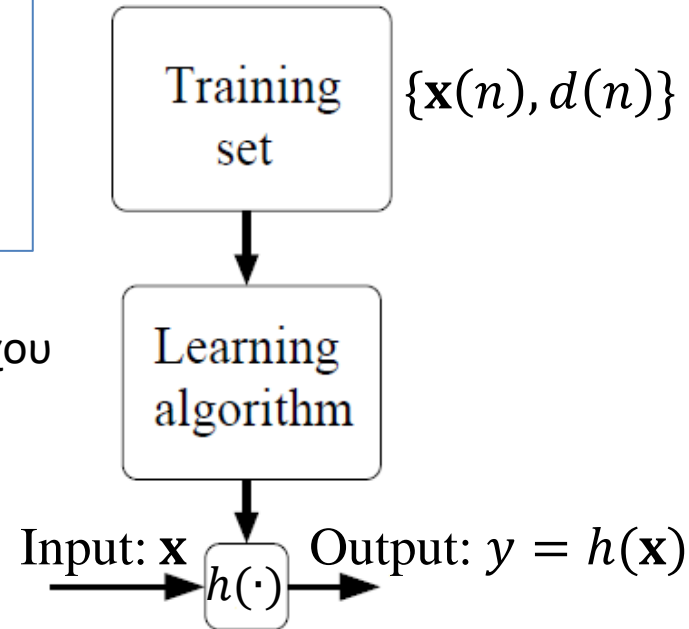
Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning (Επανάληψη)

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

- Στόχος του συστήματος είναι η αντιστοίχιση ενός δειγματικού στοιχείου εισόδου (**input sample point, example**) $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ σε τιμές εξόδου y που εκτιμούν επιθυμητές (**desired**) τιμές d (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 της y πρόκειται για πρόβλημα Ταξινόμησης, **Classification** (για δύο επιλογές έχουμε δυαδική ταξινόμηση)
- Αν η έξοδος y λαμβάνει συνεχείς τιμές, το πρόβλημα αναφέρεται σαν Παλινδρόμηση, **Regression**

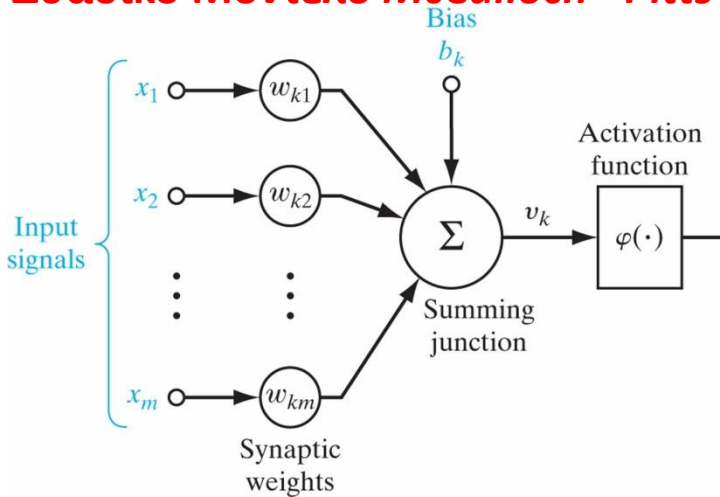


ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Επιβλεπόμενη Μάθηση & Νευρωνικά Δίκτυα (Επανάληψη)

Διαδικό Μοντέλο McCulloch - Pitts

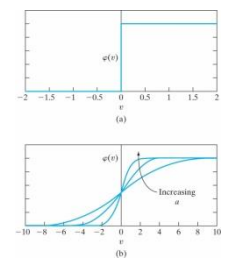
Κατάσταση Νευρώνα $k \equiv$ Διαδική έξοδος y_k
 $y_k = \varphi(v_k) \in \{-1(\text{Inactive}), +1(\text{Active})\}$



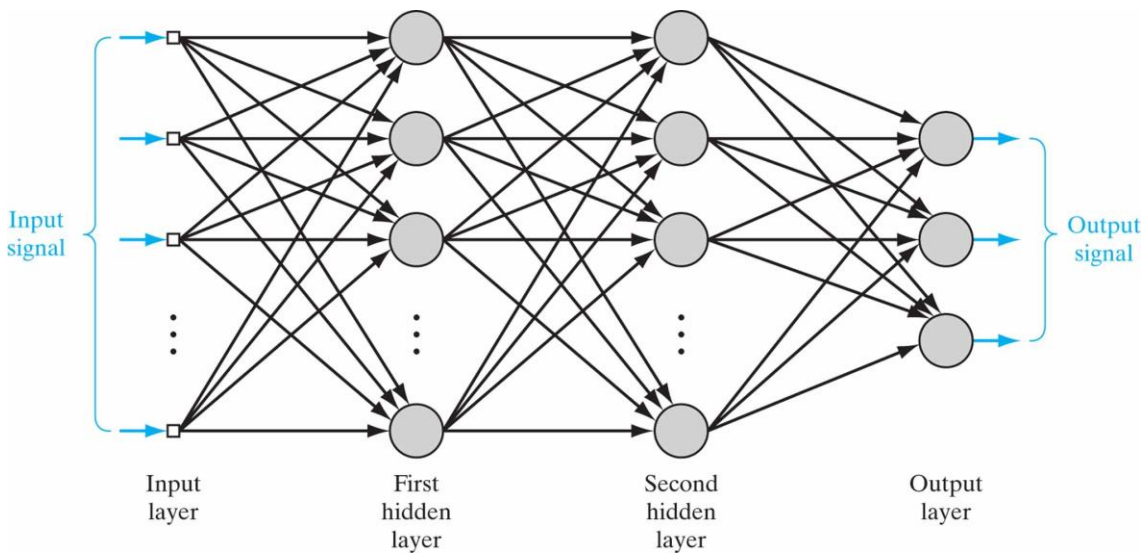
• **Signum Function:** $\varphi(v_k) = \text{sgn}(v_k) = \begin{cases} -1, & v_k < 0 \\ 0, & v_k = 0 \\ +1, & v_k > 0 \end{cases}$

Εναλλακτικές επιλογές $\varphi(\cdot)$:

- **Threshold Function:** $\varphi(v_k) = \begin{cases} 0, & v_k < 0 \\ 1, & v_k \geq 0 \end{cases}$
- **Hyperbolic Tangent:** $\varphi(v_k) = \tanh(v_k)$
- **Logistic Function:** $\varphi(v_k) = \frac{1}{1 + \exp(-av_k)}$
- **Rectified Linear Units ReLU:** $\varphi(v_k) = \max\{0, v_k\}$



Multilayer Perceptron (MLP)



Επιβλεπόμενη Μάθηση

Back Propagation Algorithm

Προσδιορισμός βαρών w_k ώστε να ελαχιστοποιείται η μέση απόκλιση (**MSE**) ορατών εξόδων $y_k(n)$ από τα **labels (targets)** $d(n)$:

$$\min_{w_k} \left\{ \frac{1}{2N} \sum_{n=1}^N [d(n) - y_k(n)]^2 \right\}$$

διαδοχικά, με εισόδους των N παραδειγμάτων μάθησης $x(n)$ σε δύο φάσεις: **Forward, Backward**

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μη Επιβλεπόμενη Μάθηση – *Unsupervised Learning, Hebbian Learning*

- Εκτίμηση *a-priori* πιθανοτήτων $p(\mathbf{x})$ του δειγματικού στοιχείου $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ με m χαρακτηριστικά - *features* x_i π.χ. *K-Means Clustering* - επιλογή K Κέντρων Βάρους Συστοιχιών (*cluster centroids*) και κατανομή σε *clusters* των σημείων \mathbf{x}
- Δεν υπάρχει εκ των προτέρων κατηγοροποίηση (*labelling*) δεδομένων. Το σύστημα με βάση στατιστικές εκτιμήσεις για τα χαρακτηριστικά (*features*) των δειγματικών στοιχείων μάθησης (*training examples*) \mathbf{x} και σχεδιαστικούς κανόνες (π.χ. κανόνα μάθησης *Hebb*) ορίζει συνάρτηση εξόδου $y = h_{\mathbf{w}}(\mathbf{x})$. Είσοδοι νέων δειγματικών στοιχείων *test* δοκιμάζουν την ικανότητα γενίκευσης (*generalization*) του μοντέλου $h_{\mathbf{w}}(\cdot)$ που βασίστηκε σε στατιστικές παραδοχές του δείγματος μάθησης (*training dataset*) & επικύρωσης (*validation dataset*)
- Η μη επιβλεπόμενη μάθηση περιλαμβάνει συστήματα αυτο-οργάνωσης (π.χ. *Self-Organizing Maps - SOM, Autoencoders*) και μαθηματικά εργαλεία επιλογής κυριάρχων χαρακτηριστικών (*principal components*) για την αποτελεσματική επεξεργασία - αποθήκευση - ταξινόμηση δεδομένων, π.χ. για επεξεργασία σήματος (*speech - image processing*) και αναγνώριση προτύπων (*pattern recognition*)

Διευκρίνιση: Στατιστική Θεώρηση Δειγματικών Δεδομένων

Ένα δείγμα (*sample*) δεδομένων είναι (υπο)σύνολο N διανυσματικών δειγματικών στοιχείων (παραδειγμάτων) \mathbf{x} που επιλέγονται ανάμεσα στα διανύσματα του δειγματικού χώρου. Οι συνιστώσες x_i των διανυσμάτων $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ κωδικοποιούν τα m χαρακτηριστικά (*features*) τους σαν δειγματικές τιμές (*sample values*) τυχαίων μεταβλητών

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διαμόρφωση Συστάδων μέσω K - Means Clustering

Οργάνωση σε K συστάδες (*clusters*) N δειγματικών στοιχείων $\mathbf{x}(n) = [x_1(n) \ x_2(n) \ \dots \ x_m(n)]^T$ με βάση κοινά χαρακτηριστικά από *unlabeled* δείγμα χωρίς δάσκαλο (*unsupervised learning*)

- **Καθορισμός Encoder** $C(n) = j$: Το $\mathbf{x}(n)$, $n = 1, 2, \dots, N$ ανήκει στο *cluster* $j = 1, 2, \dots, K$
- **Symmetric Measure of Similarity** $d(\mathbf{x}(n), \mathbf{x}(n')) = d(\mathbf{x}(n'), \mathbf{x}(n))$

Τετραγωνική Ευκλείδεια Απόσταση: $d(\mathbf{x}(n), \mathbf{x}(n')) \triangleq \|\mathbf{x}(n) - \mathbf{x}(n')\|^2$

- **Εκτιμώμενο Κέντρο Βάρους** $\hat{\boldsymbol{\mu}}_j$ (*centroid*) του *cluster* $j = 1, 2, \dots, K$ (μέσες Ευκλείδειες αποστάσεις των $\mathbf{x}(i)$ από $\hat{\boldsymbol{\mu}}_j$ για τις επιλογές του encoder $C(n) = j$)

- **Συνάρτηση Κόστους**

$$J(C) = \frac{1}{2} \sum_{j=1}^K \sum_{C(n)=j} \sum_{C(n')=j} \|\mathbf{x}(n) - \mathbf{x}(n')\|^2 = \sum_{j=1}^K \sum_{C(n)=j} \|\mathbf{x}(n) - \hat{\boldsymbol{\mu}}_j\|^2$$

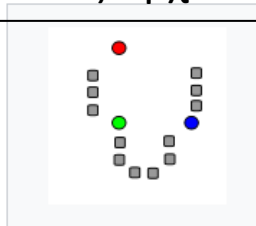
- **Κριτήριο Ελαχιστοποίησης**: Διαφορά $\hat{\boldsymbol{\sigma}}_j^2 \triangleq \sum_{C(n)=j} \|\mathbf{x}(n) - \hat{\boldsymbol{\mu}}_j\|^2$, $\min_C J(C) = \min_C \sum_{j=1}^K \hat{\boldsymbol{\sigma}}_j^2$

Σύνοψη Αλγορίθμου

- Αρχική επιλογή *centroids*: Αυθαίρετες (τυχαίες) τοποθετήσεις K σημείων
- Τοποθέτηση των N σημείων $\mathbf{x}(n)$ στο πλησιέστερο *centroid*, ανανέωση εκτιμήσεων *centroids* $\hat{\boldsymbol{\mu}}_j, j = 1, 2, \dots, K$ και επαναλήψεις μέχρι τη τελική σύγκληση του $C(n) = j$
- Η εκτέλεση του αλγορίθμου είναι αποτελεσματική αλλά χωρίς εγγύηση βέλτιστης λύσης (ανάγκη επαναλήψεων με διαφορετικές αρχικές επιλογές)

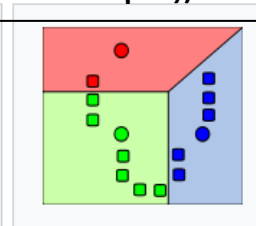
Παράδειγμα: $K = 3, N = 12$

(https://en.wikipedia.org/wiki/K-means_clustering)



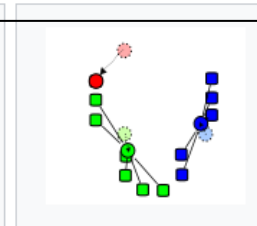
Αρχικοποίηση

Centroids



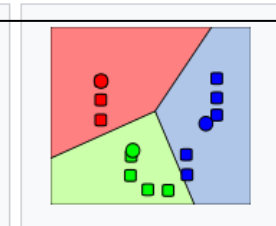
1^η Διαμόρφωση

Clusters



Προσδιορισμός

Νέων *Centroids*



Τελική Πρόταση

Clusters

The Curse of Dimensionality:

Σε ένα δειγματικό χώρο δεδομένων $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ η κωδικοποίηση των χαρακτηριστικών του δείγματος μπορεί να απαιτεί μεγάλο αριθμό διακριτών στοιχείων x_i , $i = 1, 2, \dots, m$. Για καταγραφή με στατιστική επάρκεια όλων των χαρακτηριστικών απαιτείται αριθμός δειγματικών στοιχείων N πολλαπλάσιος του m (π.χ. $N \gg 5m$, https://en.wikipedia.org/wiki/Curse_of_dimensionality)

Reduction of Dimensionality - Principal Components:

Με εφαρμογή *Unsupervised Learning* σε γραμμικό σύστημα με είσοδο $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ μπορεί να μειωθεί ο αριθμός των χαρακτηριστικών (αριθμός συντεταγμένων m) μέσω μετασχηματισμού σε **Ασυσχέτιστες Κύριες Συνιστώσες** (*Principal Components*) και έξοδο $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_l]^T$ με επιλογή των σημαντικότερων συνιστωσών ($l \ll m$ χαρακτηριστικών) με τη μεγαλύτερη αναμενόμενη διασπορά

Μεθοδολογίες για Επιλογή Κυρίων Συνιστωσών:

- **Covariance Method**: Μέσω στατιστικής ανάλυσης του διανυσματικού δειγματικού χώρου μάθησης, μετασχηματισμού σε ορθοκανονικό (*orthonormal*) σύστημα συντεταγμένων και αλγορίθμων επιλογής σημαντικών (*principal*) συνιστωσών (αναλογία με μεθόδους *Γραμμικής Άλγεβρας* και *Θεωρίας Επικοινωνιών* - μετασχηματισμοί *Karhunen - Loève*)
- **Hebbian Learning Method**: Μέσω αυτοοργάνωσης νευρωνικών δικτύων με τοπικές ρυθμίσεις κανόνων *Hebb* σε μη επιβλεπόμενη μάθηση (*unsupervised learning*)

Orthonormal Transformation σε Principal Components (1/3)

Ορισμοί Στατιστικής Προσέγγισης (Covariance Method) της PCA

- Θεωρούμε είσοδο από δειγματικά στοιχεία δεδομένων $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ με m χαρακτηριστικά (*features*) κωδικοποιημένα στις συνιστώσες x_i
- Οι συνιστώσες x_i αποτελούν *δειγματικές τιμές* (sample values) *τυχαίων μεταβλητών* X_i που αναφέρονται σε τυχαία διανύσματα $\mathbf{X} = [X_1 \ X_2 \ \dots \ X_m]^T$ του δειγματικού χώρου. Υποθέτουμε πως $E[\mathbf{X}] = E[X_i] = 0$
- Η συμμετρική μήτρα ($m \times m$) $\mathbf{R} = E[\mathbf{X} \mathbf{X}^T]$ είναι η μήτρα συσχέτισης (*Correlation Matrix*) των τυχαίων διανυσμάτων \mathbf{X} με ιδιοδιανύσματα (*eigenvectors*) \mathbf{q}_j και ιδιοτιμές (*eigenvalues*) λ_j : $\mathbf{R} \mathbf{q}_j = \lambda_j \mathbf{q}_j$, $j = 1, 2, \dots, m$ αριθμημένα κατά φθίνουσα σειρά των ιδιοτιμών:
$$\lambda_j \mathbf{q}_j^T \mathbf{q}_k = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \quad \text{και} \quad \mathbf{q}_j^T \mathbf{R} \mathbf{q}_k = \begin{cases} \lambda_j, & k = j \\ 0, & k \neq j \end{cases}$$
- Τα ιδιοδιανύσματα \mathbf{q}_j ορίζουν ορθοκανονικές (*orthonormal*) κύριες (*principal*) κατευθύνσεις που μετασχηματίζουν το στοιχείο $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ με m συντεταγμένες x_i σε $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_m]^T = [\mathbf{x}^T \mathbf{q}_1 \ \mathbf{x}^T \mathbf{q}_2 \ \dots \ \mathbf{x}^T \mathbf{q}_m]$ με m συντεταγμένες a_i που αποτελούν τις **Κύριες Συνιστώσες** (*Principal Components*). Η αρίθμηση ακολουθεί τη φθίνουσα σειρά των $\lambda_j = \mathbf{q}_j^T \mathbf{R} \mathbf{q}_j = \text{var}(A_j) \triangleq \sigma_j^2$ όπου A_j τυχαία μεταβλητή με δειγματική τιμή την κύρια συνιστώσα a_j
- Οι αρχικές συντεταγμένες προκύπτουν μονοσήμαντα από τις Κύριες Συνιστώσες:

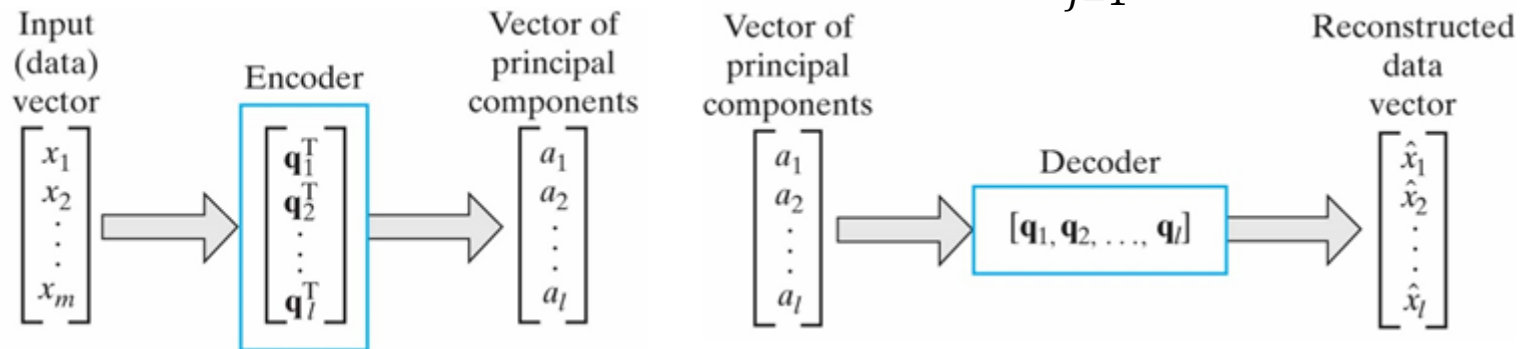
$$\mathbf{x} = \sum_{j=1}^m a_j \mathbf{q}_j$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Orthonormal Transformation σε Principal Components (2/3)

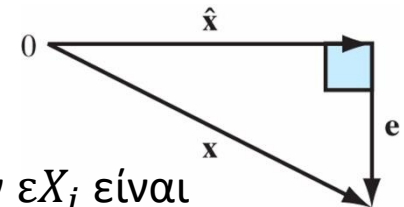
Αν αγνοήσουμε τις κύριες συνιστώσες (*principal components*) με τις μικρότερες διασπορές $\sigma_j^2 = \lambda_j$, $j = l + 1, l + 2, \dots, m$ προκύπτει προσεγγιστική εκτίμηση του στοιχείου \mathbf{x} από το στοιχείο $\hat{\mathbf{x}}$ με μικρότερο αριθμό συνιστωσών $l < m$

$$\hat{\mathbf{x}} = [\hat{x}_1 \hat{x}_2 \dots \hat{x}_l] = [\mathbf{q}_1 \mathbf{q}_2 \dots \mathbf{q}_l][a_1 a_2 \dots a_l]^T = \sum_{j=1}^l a_j \mathbf{q}_j \text{ για } l < m$$



Το σφάλμα εκτίμησης $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \sum_{i=l+1}^m a_i \mathbf{q}_i$ είναι διάνυσμα ορθογώνιο προς το $\hat{\mathbf{x}}$:

$$\mathbf{e}^T \hat{\mathbf{x}} = \sum_{i=l+1}^m a_i \mathbf{q}_i^T \sum_{j=1}^l a_j \mathbf{q}_j = 0$$



- Η συνολική διασπορά των m ανεξαρτήτων τυχαίων μεταβλητών εX_j είναι

$$\sum_{j=1}^m \sigma_j^2 = \sum_{j=1}^m \lambda_j$$

- Η συνολική διασπορά των l τυχαίων μεταβλητών A_j είναι $\sum_{j=1}^l \sigma_j^2 = \sum_{j=1}^l \lambda_j$

\Rightarrow Η συνολική διασπορά του σφάλματος $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ είναι $\lambda_{l+1} + \lambda_{l+2} + \dots + \lambda_m$ που αντιστοιχεί στις κύριες συνιστώσες με τις μικρότερες διακυμάνσεις

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Orthonormal Transformation σε Principal Components (3/3)

Εφαρμογή PCA για Ψηφιακή Συμπύεση & Αναγνώριση Εικόνων Χειρόγραφων Αριθμών
Image Compression & Pattern Recognition of Handwritten Numbers

Δειγματικά Στοιχεία Μάθησης:

Σκαναρισμένες εικόνες δέκα χειρόγραφων αριθμών $\{0, 1, \dots, 9\}$

$N = 1700$ στοιχεία/αριθμό



Στήλη 1: Κωδικοποίηση με $m = 32 \times 32 = 1024$ δυαδικά ψηφία

Στήλη 2: Δειγματικοί Μέσοι Όροι για κανονικοποίηση

Υπολογισμός των $l = 64$ σημαντικότερων **ιδιοδιανυσμάτων** της **Μήτρας Συσχέτισης** $m \times m = (1024) \times (1024)$ μετά από κανονικοποίηση των δειγματικών στοιχείων (αφαίρεση δειγματικών μέσων όρων)

Ανασύσταση Εικόνας με $l \leq 64$ Principal Components

$$l \ll m = 32 \times 32 = 1024$$

Στήλη 3: $l = 1$

Στήλη 4: $l = 5$

Στήλη 5: $l = 16$

Στήλη 6: $l = 32$ (ικανοποιητική διάκριση αριθμών)

Στήλη 7: $l = 64$ (άψογη διάκριση αριθμών με μεγάλη συμπίεση)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός 1^{ης} Κύριας Συνιστώσας από Γραμμικό Νευρώνα με Μάθηση Hebbian (1/2)

Αρχές Αυτο-οργάνωσης Χαρακτηριστικών (Self-organized Feature Analysis)

Αξίωμα του Hebb: Αν τα σήματα στα άκρα μιας νευρωνικής σύναψης i ενεργοποιούνται ταυτόχρονα (*synchronous update*) που ορίζεται από το βήμα n , το συναπτικό της βάρος $w_i(n)$ ενισχύεται. Αλλιώς φθίνει προς μηδενισμό (*παραλλαγή από αρχικό αξίωμα για μάθηση με όρους νευροψυχολογίας*)

Αρχή Ανταγωνιστικότητας: Οι πιο ενεργές συνάψεις τείνουν να μηδενίσουν τις λιγότερο ενεργές

Νευρωνικό Δίκτυο Υπολογισμού *First Principal Component* Δείγματος με m Features **Hebbian-based Maximum Eigenfilter**

$$y(n) = \sum_{i=1}^m w_i(n)x_i(n) \text{ στην επανάληψη } n$$

Τα βάρη αυξάνονται στην επανάληψη $n \rightarrow n + 1$ αν $y(n)x_i(n) > 0$ (*αξίωμα Hebb*)

$$w_i(n + 1) = w_i(n) + \eta y(n)x_i(n), \quad i = 1, 2, \dots, m \quad \text{όπου } \eta \text{ learning (hyper)parameter}$$

Απαιτείται **Normalization** για ευστάθεια. Με βάση την **Αρχή της Ανταγωνιστικότητας** διαιρούμε σε κάθε βήμα με το εκτόπισμα των $w_k(n)$ που μειώνουν την αύξηση εις βάρος τους:

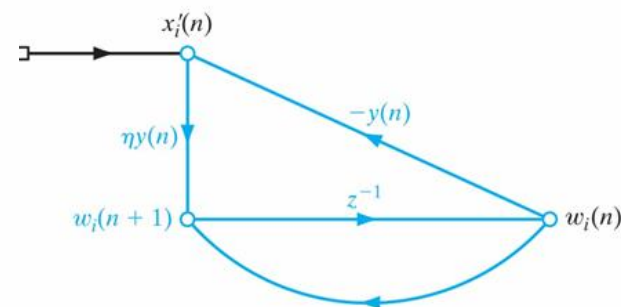
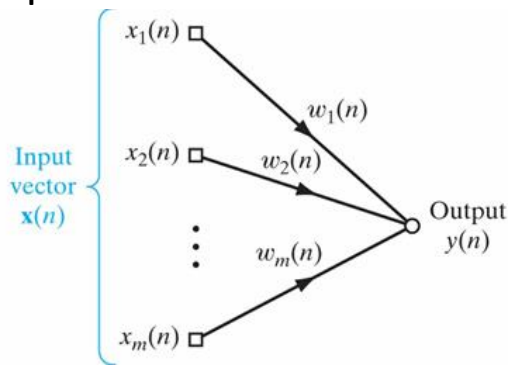
$$w_i(n + 1) = \frac{w_i(n) + \eta y(n)x_i(n)}{(\sum_{k=1}^m [w_k(n) + \eta y(n)x_k(n)]^2)^{1/2}} \cong w_i(n) + \eta y(n)[x_i(n) - y(n)w_i(n)]$$

Η προσέγγιση ισχύει για μικρές τιμές της παραμέτρου η

Η θετική ανάδραση $y(n)x_i(n)$ που προκαλεί αστάθεια αντισταθμίζεται από την $-y(n)w_i(n)$:

$$w_i(n + 1) = w_i(n) + \eta y(n)x'_i(n)$$

$$x'_i(n) = x_i(n) - y(n)w_i(n)$$



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός 1^{ης} Κύριας Συνιστώσας από Γραμμικό Νευρώνα με Μάθηση Hebbian (2/2)

Ζητήματα Σύγκλισης

Διανυσματικοί ορισμοί

$$\mathbf{x}(n) = [x_1(n) \ x_2(n) \ \dots \ x_m(n)]^T$$

$$\mathbf{w}(n) = [w_1(n) \ w_2(n) \ \dots \ w_m(n)]^T$$

$$y(n) = \mathbf{w}^T(n)\mathbf{x}(n), \quad \mathbf{w}(n+1) = \mathbf{w}(n) + \eta y(n)[\mathbf{x}(n) - y(n)\mathbf{w}(n)] \Rightarrow$$

Αλγόριθμος αυτό-οργανωμένης μη επιβλεπόμενης μάθησης:

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

- Οι όροι $\mathbf{x}(n)\mathbf{x}^T(n)$ αποτελούν μια στιγμιαία υλοποίηση της μήτρας συσχέτισης $\mathbf{R} = E[\mathbf{X}\mathbf{X}^T]$ χωρίς μέσους όρους και οδηγούν προς την ανάλυση της σύγκλισης του αλγορίθμου μέσω μη γραμμικής στοχαστικής εξίσωσης διαφορών (εκτός ορίων του μαθήματος)
- Δεν υπάρχει εξωτερική επιρροή στον αλγόριθμο μάθησης λόγω της μη επιβλεπόμενης (αυτό-οργανωμένης, *self-organized*) φύσης του πλην του *a-priori* καθορισμού της (*hyper*)*parameter* μάθησης η

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

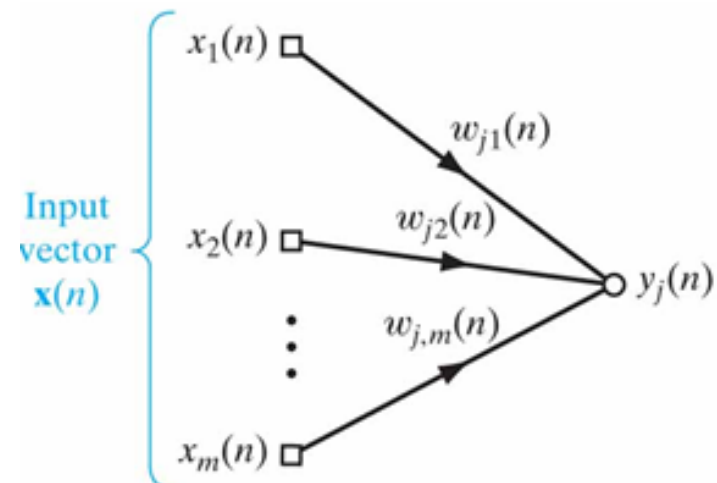
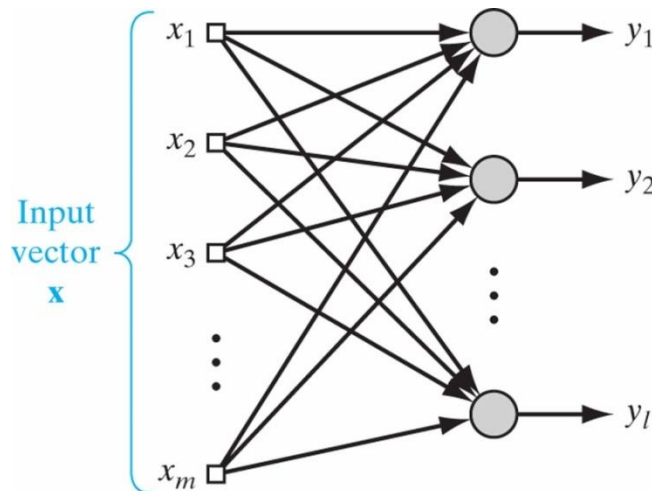
Προσδιορισμός l Κυρίων Συνιστωσών από Γραμμικούς Νευρώνες με Μάθηση Hebbian (1/3) (Hebbian-based Principal-Component Analysis)

Γενίκευση του Hebbian-based maximum Eigenfilter: Γραμμικό Feedforward Νευρωνικό Δίκτυο ενός επιπέδου με l νευρώνες για προσδιορισμό των l σημαντικών Principal Components από Δεδομένα Διαστάσεων m , $l < m$:

$$y_j(n) = \sum_{i=1}^m w_{ji}(n)x_i(n), \quad j = 1, 2, \dots, l$$

Τα βάρη $w_{ji}(n)$ από την είσοδο $x_i(n)$, $i = 1, 2, \dots, m$ προς την j κύρια συνιστώσα $y_j(n)$, $j = 1, 2, \dots, l$ διαφοροποιούνται κατά $\Delta w_{ji}(n)$ στην επανάληψη $n \rightarrow n + 1$:

$$\Delta w_{ji}(n) = \eta \left[y_j(n)x_i(n) - y_j(n) \sum_{k=1}^j w_{ki}(n)y_k(n) \right], \quad \text{όπου } j = 1, 2, \dots, l \text{ \& } i = 1, 2, \dots, m$$



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός l Κυρίων Συνιστωσών από Γραμμικούς Νευρώνες με Μάθηση Hebbian (2/3) (Hebbian-based Principal-Component Analysis)

Γενικευμένος Αλγόριθμος του Hebb – Generalized Hebbian Algorithm (GHA)

$$\Delta w_{ji}(n) = \eta y_j(n) [x'_i(n) - w_{ji}(n) y_j(n)], \quad j = 1, 2, \dots, l \text{ και } i = 1, 2, \dots, m$$

$$x'_i(n) = x_i(n) - \sum_{k=1}^{j-1} w_{ki}(n) y_k(n)$$

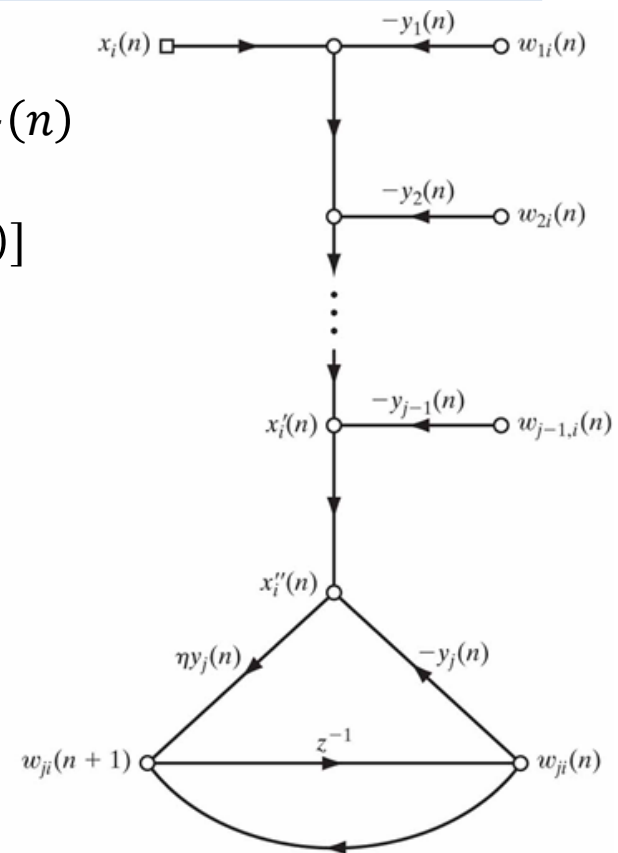
$$\Delta w_{ji}(n) = \eta y_j(n) x''_i(n) \text{ όπου } x''_i(n) = x'_i(n) - w_{ji}(n) y_j(n)$$

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n), w_{ji}(n) = z^{-1} [w_{ji}(n+1)]$$

Διανυσματική Μορφή:

$$\Delta \mathbf{w}_j(n) = \eta y_j(n) \mathbf{x}'_i(n) - \eta y_j^2(n) \mathbf{w}_j(n), j = 1, 2, \dots, l$$

$$\text{όπου } \mathbf{x}'(n) = \mathbf{x}(n) - \sum_{k=1}^{j-1} \mathbf{w}_k(n) y_k(n)$$



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός l Κυρίων Συνιστωσών από Γραμμικούς Νευρώνες με Μάθηση Hebbian (3/3) (Hebbian-based Principal-Component Analysis)

$$\mathbf{x}'(n) = \mathbf{x}(n) - \sum_{k=1}^{j-1} \mathbf{w}_k(n) y_k(n), \quad j = 1, 2, \dots, l$$

Για $j = 1$: $\mathbf{x}'(n) = \mathbf{x}(n)$

Προσδιορισμός 1^{ης} κύριας συνιστώσας $y_1(n)$

Για $j = 2$: $\mathbf{x}'(n) = \mathbf{x}(n) - \mathbf{w}_1(n) y_1(n)$

Προσδιορισμός 2^{ης} συνιστώσας $y_2(n)$ σαν 1^η συνιστώσα μετά από αφαίρεση της ήδη υπολογισθείσας $y_1(n)$

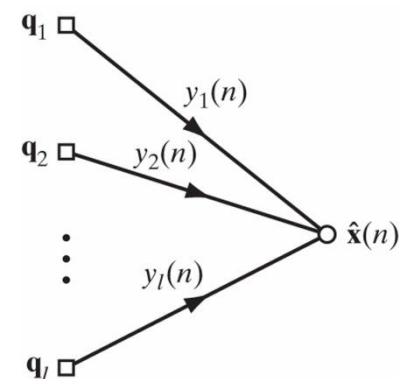
Για $j = 3$: $\mathbf{x}'(n) = \mathbf{x}(n) - \mathbf{w}_1(n) y_1(n) - \mathbf{w}_2(n) y_2(n)$

Προσδιορισμός 3^{ης} συνιστώσας $y_3(n)$ σαν 1^η συνιστώσα μετά από αφαίρεση των ήδη υπολογισθείσας $y_1(n)$ και $y_2(n)$

.....

Οι l πιο σημαντικές κύριες συνιστώσες αντιστοιχούν στα ιδιοδιανύσματα \mathbf{q}_k της μήτρας συσχέτισης $\mathbf{R} = E[\mathbf{X}\mathbf{X}^T]$, $k = 1, 2, \dots, l$ αριθμημένες κατά τη φθίνουσα σειρά των ιδιοτιμών $\lambda_1 > \lambda_2 > \dots > \lambda_l$ και δίνουν την εκτίμηση $\hat{\mathbf{x}}(n)$ με l χαρακτηριστικά του δειγματικού στοιχείου $\mathbf{x}(n)$ με $m > l$

$$\hat{\mathbf{x}}(n) = \sum_{k=1}^l y_k(n) \mathbf{q}_k \quad \text{για } l < m$$



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Εφαρμογή PCA μέσω Generalized Hebbian Algorithm (GHA) στη Κωδικοποίηση Εικόνας (1/2)

- Αρχική Εικόνα (Δείγμα) **Lena**: 256×256 pixels, 256 gray levels στην **1^η εικόνα**
- Μάθηση μέσω $N = 2000$ σκαναρισμένων εικόνων, χωρισμένων σε 1024 μη επικαλυπτόμενα τμήματα (**Blocks**) μεγέθους 8×8 pixels: $m = 64$ features/block
- Το κάθε block ορίζει δειγματικό στοιχείο εισόδου σαν διάνυσμα από m features (pixels) κωδικοποιημένων σε 256 gray levels (8 bits/pixel):

$$\mathbf{x}(n) = [x_1(n) \ x_2(n) \ \dots \ x_m(n)]^T \quad n = 1, 2, \dots, N$$

- Το δείγμα τροφοδοτεί **Μονοστρωματικό Linear Feedforward Δίκτυο** με $l = 8$ Νευρώνες εξόδου
- Το σύστημα συγκλίνει σε $m \times l = 64 \times 8$ συναπτικά βάρη $w_{ji}(n)$ και απολήγει σε 8 εξόδους (κυριότερες συνιστώσες, **Significant Principal Components**):

$$y_j(n) = \sum_{i=1}^m w_{ji}(n)x_i(n), \quad j = 1, 2, \dots, l$$

- Ρυθμός μάθησης (**Learning Rate**): $\eta = 10^{-4}$
- Τα βάρη (**weights**) μετά τη σύγκλιση παρίστανται στην **2^η εικόνα** με $4 \times 2 = 8$ περιοχές (**masks**), 64 τμήματα/mask, σύνολο $64 \times 8 = 1024$ τμήματα που απεικονίζουν την συνεισφορά των 64 χαρακτηριστικών του δείγματος εισόδου στις 8 εξόδους. Το άσπρο χρώμα υποδηλώνει θετική συνεισφορά ενός χαρακτηριστικού, το μαύρο αρνητική και το γκρι μηδενική
- Στη **3^η εικόνα** παρουσιάζεται ανασύσταση της αρχικής με τη συμμετοχή μόνο των $l = 8$ κυριότερων συνιστωσών της (**l Principal Components**) χωρίς κβαντισμό

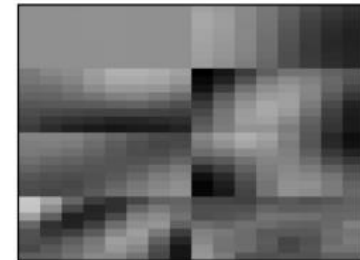
$$\hat{\mathbf{x}}(n) = \sum_{k=1}^l y_k(n) \mathbf{q}_k, \quad \mathbf{q}_k = \lim_n \mathbf{w}_k(n), \quad \mathbf{w}_k(n) = [w_{k1}(n) \ w_{k2}(n) \ \dots \ w_{km}(n)]^T$$

- Η **4^η εικόνα** αποτελεί συμπίεση της **3^{ης} εικόνας** με κβαντισμένες τιμές των βαρών ανάλογα με το λογάριθμο της διασποράς τους (~ 0.53 bits/pixel)

Original Image



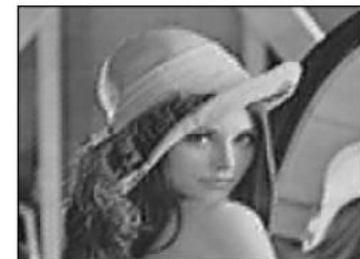
Weights



Using First 8 Components



11 to 1 compression



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

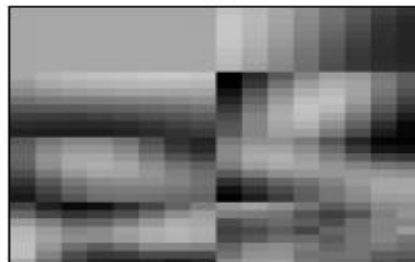
Εφαρμογή PCA μέσω Generalized Hebbian Algorithm (GHA) στη Κωδικοποίηση Εικόνας (2/2)

- Αρχική Εικόνα (**Peppers**): 256×256 pixels, 256 gray levels στην *πάνω εικόνα*

Original Image



Weights



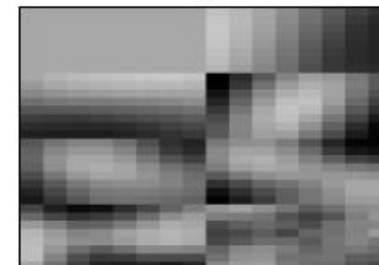
Using First 8 Components



12 to 1 compression μέσω κβαντισμού βαρών των 8 κυριότερων συνιστωσών, με βάρη όπως προσδιορίστηκαν για την εικόνα **Peppers**



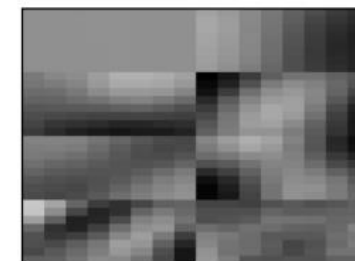
Weights



12 to 1 compression μέσω κβαντισμού βαρών των 8 κυριότερων συνιστωσών, με βάρη όπως προσδιορίστηκαν για την εικόνα **Lena** αλλά εφαρμόστηκαν στην εικόνα **Peppers** (**GENERALIZATION ?**)



Weights



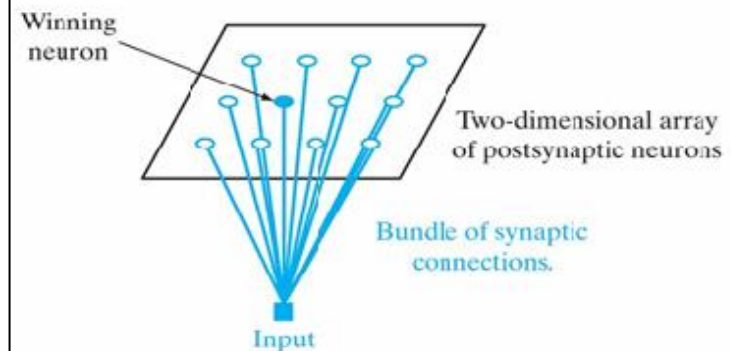
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Απεικόνιση Διανυσματικών Δειγματικών Στοιχείων σε Νευρωνικά Δίκτυα με Τοπολογικά Κριτήρια Αυτοοργάνωσης, **Self-Organizing Maps (SOM)**

- Τα **SOM** είναι Νευρωνικά Δίκτυα με εισόδους δειγματικά στοιχεία (**παραδείγματα**) $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ διανυσματικού δειγματικού χώρου m διαστάσεων (**χαρακτηριστικών**)
- Σε αναλογία με τρόπους αποθήκευσης και επεξεργασίας στον εγκέφαλο των θηλαστικών, ο **Kohonen** πρότεινε το 1982 μοντέλο **μονοστρωματικού Feedforward Neural Network** με μη γραμμικούς νευρώνες $j = 1, 2, \dots, l$ σε δυσδιάστατο πλέγμα (**lattice**) και με κατευθυντικές συνάψεις $\mathbf{w}_j = [w_{j1} \ w_{j2} \ \dots \ w_{jm}]^T$ από τους κόμβους εισόδου προς τους (**postsynaptic**) νευρώνες του πλέγματος
- Οι έξοδοι των **postsynaptic** νευρώνων, εφόσον ενεργοποιηθούν, συνθέτουν την εκτίμηση του νευρωνικού συστήματος για το παράδειγμα (διάνυσμα) εισόδου $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ με προσέγγιση σε πρότυπα (**patterns**) που αποθηκεύτηκαν σε περιοχές του πλέγματος. Οι περιοχές αντιστοιχούν σε γειτονιές ενεργών νευρώνων, που καθορίζονται από τον πλησιέστερο νευρώνα (**winner**) προς το διάνυσμα \mathbf{x} με μη επιβλεπόμενη **Ανταγωνιστική Μάθηση (Competitive Learning)** και διαδικασία αυτοοργάνωσης (**Self-Organizing Map**)

Εφαρμογές SOM

- Επιλογή κυρίαρχων χαρακτηριστικών σε πολυδιάστατα δείγματα
- Συμπύεση εικόνων με εντοπισμό παρομοίων υποπεριοχών
- Αναγνώριση προτύπων, ταξινόμηση παραδειγμάτων
- Φιλτάρισμα από παρεμβολές - θορύβους
- Συμπλήρωση ατελών παραδειγμάτων



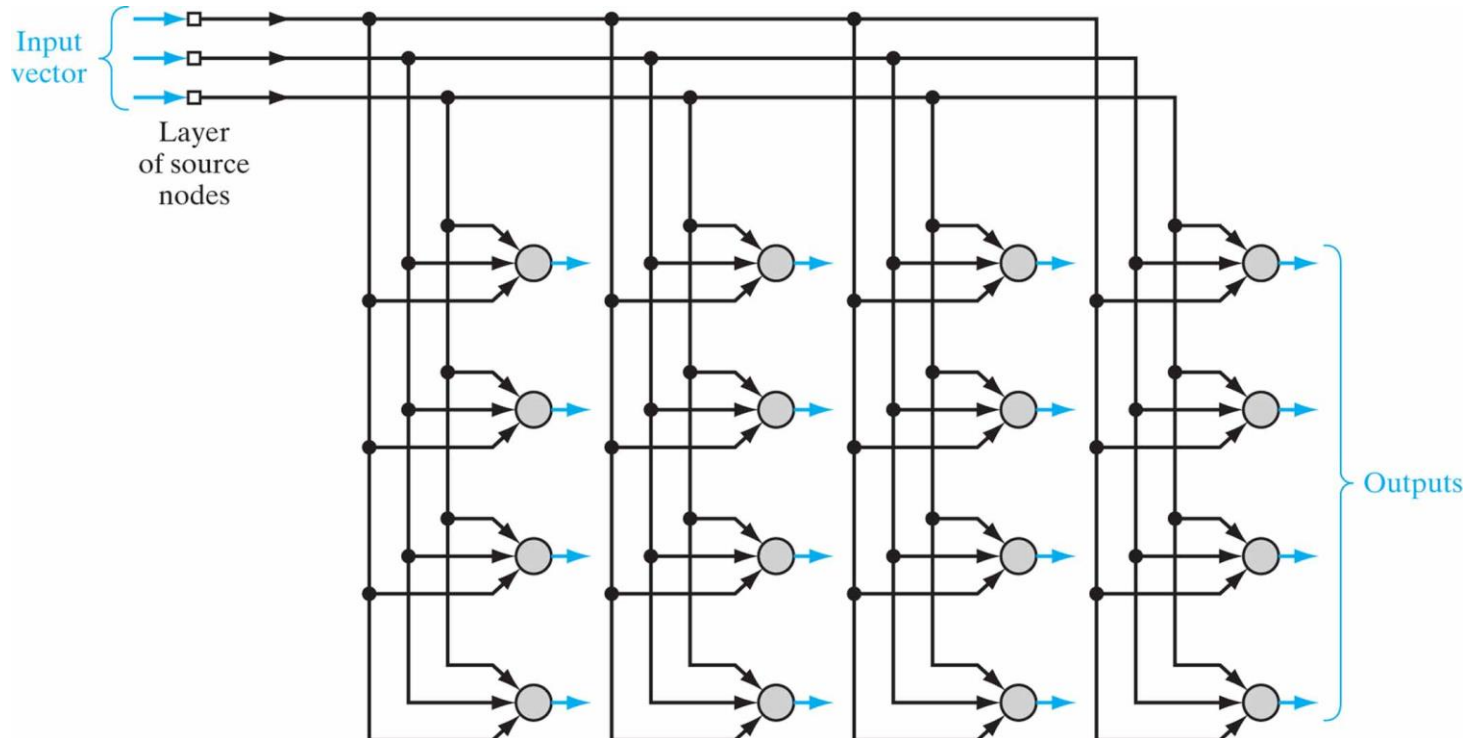
Μοντέλο SOM, **Kohonen** 1982

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Self-Organizing Maps (SOM) (1/5)

- Ο αλγόριθμος διαμόρφωσης του *feature map* συνδυάζει τις αρχές νευροφυσιολογίας του *Hebb* για τοπολογική αυτοοργάνωση νευρωνικών δικτύων σε πλέγματα
- Αποθηκεύει μέσω μη-επιβλεπόμενης μάθησης *πρότυπα (patterns)* που προκύπτουν από το δείγμα μάθησης με επιλογή των σημαντικότερων χαρακτηριστικών τους $l \ll m$ (*data compression, dimensionality reduction*)
- Το δίκτυο *SOM* ανασυνθέτει κατ' εκτίμηση και αναγνωρίζει ελλειμματικά ή παραμορφωμένα νέα παραδείγματα (π.χ. λόγω θορύβου) με βάση τη στατιστική συνάφεια προς αποθηκευμένα πρότυπα

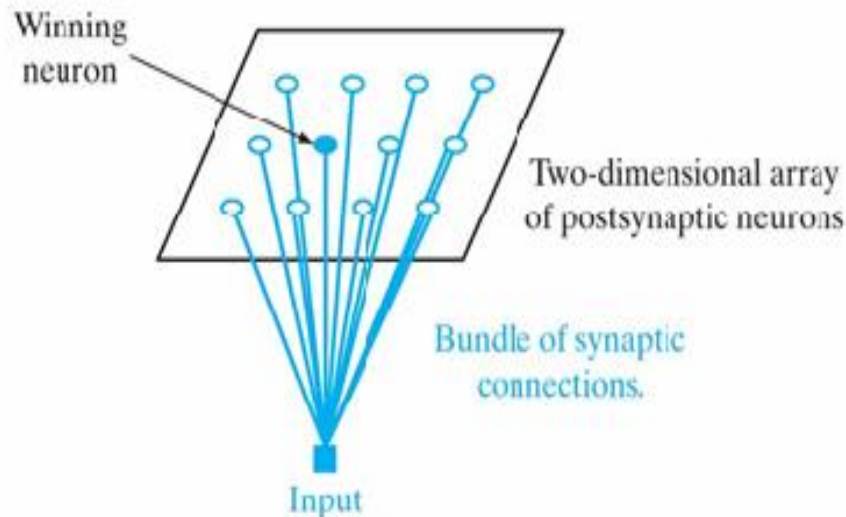
Διάταξη Πλέγματος Νευρώνων με Είσοδο 3 Χαρακτηριστικών και 4×4 Διάσταση Εξόδου



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Self-Organizing Maps (SOM) (2/5)

- Ο αλγόριθμος **μη επιβλεπόμενης μάθησης** ανακαλύπτει τον **πλησιέστερο** νευρώνα (**winning neuron**) για κάθε διάνυσμα εισόδου (δειγματικό στοιχείο μάθησης) x με **Ανταγωνιστική Διαδικασία** (**Competition Process**) που μεγιστοποιεί διακριτική συνάρτηση (**discriminant function**)
- Ο **winner** καθορίζει την περιοχή των ενεργών νευρώνων σε **Συνεργατική Διαδικασία** (**Cooperation Process**) με γειτονικούς του νευρώνες στο πλέγμα, με αποτέλεσμα την αυτό-οργάνωση ενεργών χαρακτηριστικών σε τοπογραφικό **feature map**
- Σε περιβάλλον **αυτοοργάνωσης**, που ακολουθεί τους κανόνες του **Hebb**, το διάνυσμα των παραμέτρων (συναπτικών βαρών) w_j αλλάζει με τη διαδοχική είσοδο παραδειγμάτων x . Απαιτείται διόρθωση με **Διαδικασία Προσαρμογής** (**Adaptive Process**) που να εγγυάται πως τα βάρη δεν αυξάνουν συνεχώς και επομένως ο επαναληπτικός αλγόριθμος μάθησης είναι ευσταθής



Μοντέλο SOM, **Kohonen** 1982

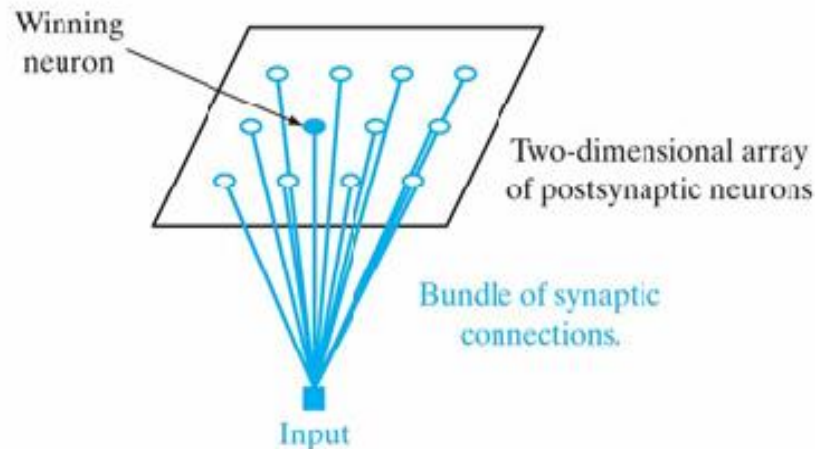
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Self-Organizing Maps (SOM) (3/5)

Ανταγωνιστική Διαδικασία

- Ο αλγόριθμος **μη επιβλεπόμενης μάθησης** ανακαλύπτει τον **πλησιέστερο** νευρώνα (**winner**) για κάθε διάνυσμα εισόδου (δειγματικό στοιχείο, παράδειγμα μάθησης) \mathbf{x} με **ανταγωνισμό** (**Competition**) που μεγιστοποιεί διακριτική συνάρτηση (**discriminant function**)
- Για ένα παράδειγμα $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$ επιλέγεται διάνυσμα συναπτικών βαρών $\mathbf{w}_j = [w_{j1} \ w_{j2} \ \dots \ w_{jm}]^T$ προς τους $j = 1, 2, \dots, l$ **postsynaptic neurons** του πλέγματος
- Επιλέγεται σαν **winning neuron** ο νευρώνας $i(\mathbf{x})$ με το μέγιστο εσωτερικό γινόμενο $\mathbf{w}_j^T \mathbf{x}$ σαν το κέντρο διέγερσης της περιοχής του πλέγματος. Η επιλογή του ισοδυναμεί με το σημείο ελάχιστης Ευκλείδειας απόστασης μεταξύ των διανυσμάτων \mathbf{x} και \mathbf{w}_j αν $\|\mathbf{w}_j\| = 1$ οπότε

$$i(\mathbf{x}) = \arg \min_j \|\mathbf{x} - \mathbf{w}_j\|$$

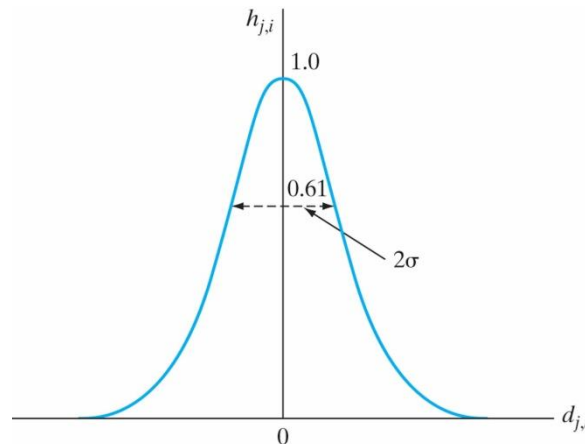


ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Self-Organizing Maps (SOM) (4/5)

Συνεργατική Διαδικασία

- Ο *winning neuron* $i(\mathbf{x})$ προσδιορίζει περιοχή $h_{j,i(\mathbf{x})}$ διεγερμένων γειτονικών νευρώνων j σε απόσταση (*lateral distance*) $d_{j,i(x)}$
- Συνήθης επιλογή: *Gaussian Function* $h_{j,i(\mathbf{x})} = \exp\left(-\frac{d_{j,i(x)}^2}{2\sigma^2}\right)$
- Η παράμετρος σ μπορεί να μειώνεται όσο προχωράει η διαδικασία μάθησης, μειώνοντας τις χωρικές εξαρτήσεις (*correlations*) των νευρώνων στο πλέγμα και τον χρόνο ανανέωσης των συναπτικών βαρών (*updates of synaptic weights*)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Self-Organizing Maps (SOM) (5/5)

Διαδικασία Προσαρμογής

- Σε περιβάλλον *αυτοοργάνωσης*, που ακολουθεί τους κανόνες του *Hebb*, το διάνυσμα των παραμέτρων (συναπτικών βαρών) \mathbf{w}_j αλλάζει με τη διαδοχική είσοδο παραδειγμάτων \mathbf{x}
- Απαιτείται διόρθωση που να εγγυάται πως τα βάρη δεν αυξάνουν συνεχώς και επομένως ο επαναληπτικός αλγόριθμος μάθησης είναι ευσταθής
- Ορισμός **όρου λήθης** (*forgetting term*) $g(y_i)\mathbf{w}_j$ όπου $g(y_i)$ θετική συνάρτηση της εξόδου y_i με $g(y_i) = 0$ για $y_i = 0$
- Οι επαναλήψεις προσδιορισμού των συναπτικών βαρών οδηγούνται από τις διαφορές
$$\Delta \mathbf{w}_j = \eta y_i \mathbf{x} - g(y_i) \mathbf{w}_j$$

η : *learning hyperparameter*,
 $y_i \mathbf{x}$: *Hebbian term*
 $g(y_i) \mathbf{w}_j$: *forgetting term*
- Με γραμμικό ορισμό του *forgetting term* $g(y_i) = \eta y_i$ και με $y_i = h_{j,i(\mathbf{x})}$ έχουμε $\Delta \mathbf{w}_j = \eta h_{j,i(\mathbf{x})} (\mathbf{x} - \mathbf{w}_j)$ όπου $i(\mathbf{x})$ ο *winning neuron* για είσοδο \mathbf{x}
- Στην επανάληψη $n \rightarrow n + 1$ με μειούμενη υπερπαραμέτρο $\eta(n)$:
$$\mathbf{w}_j(n + 1) = \mathbf{w}_j(n) + \eta(n) h_{j,i(\mathbf{x})}(n) (\mathbf{x}(n) - \mathbf{w}_j(n))$$
- Τα συναπτικά βάρη του *winning neuron* τείνουν προς το παράδειγμα εισόδου \mathbf{x} και τα συναπτικά βάρη των νευρώνων της γειτονιάς του τείνουν στατιστικά προς την κατανομή του δείγματος των \mathbf{x}

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Autoencoders

Encoder:

Input $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_m]^T$

Output $\mathbf{h} = \mathbf{F}_e(\mathbf{x}) = [h_1 \ h_2 \ \dots \ h_l]^T, l \ll m$ (*code, latent variables*)

Decoder:

Input $\mathbf{h} = [h_1 \ h_2 \ \dots \ h_l]^T$

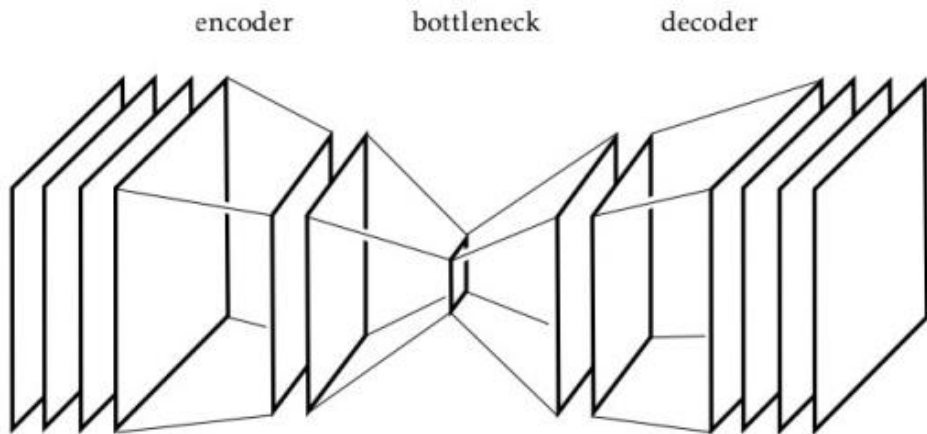
Output $\mathbf{x}' = \mathbf{F}_d(\mathbf{h}) = [x_1' \ x_2' \ \dots \ x_m']^T$ (*reconstruction* του \mathbf{x})

Bottleneck:

Ενδιάμεσα Στρώματα Κρυφών Νευρώνων που επεξεργάζονται τις l *latent variables*

Αλγόριθμος Μάθησης (*Unsupervised Learning*):

Ελαχιστοποίηση απόκλισης $\|\mathbf{x} - \mathbf{x}'\|^2$ με επαναλήψεις *backpropagation* σε εισόδους στοιχείων *unlabeled* δείγματος μάθησης



- Οι *latent variables* αποκαλύπτουν *reduced feature maps*. Αν τα νευρωνικά δίκτυα είναι γραμμικά με μηδενικά *bias* αποτελούν τις l κύριες συνιστώσες *unlabeled* δείγματος
- Μπορούν να χρησιμοποιηθούν για *image compression, pattern recognition, διόρθωση & συμπλήρωση ατελών δειγματικών στοιχείων, ανίχνευση ανωμαλιών* κλπ. όταν δεν είναι διαθέσιμο *labeled training dataset*