

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μη Επιβλεπόμενη Μάθηση - Unsupervised Learning

***K*-Means Clustering**

Ανάλυση Κυρίων Συνιστωσών - Principal Component Analysis (PCA)

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Video Conference μέσω MS Teams

Πέμπτη 19/3/2020

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

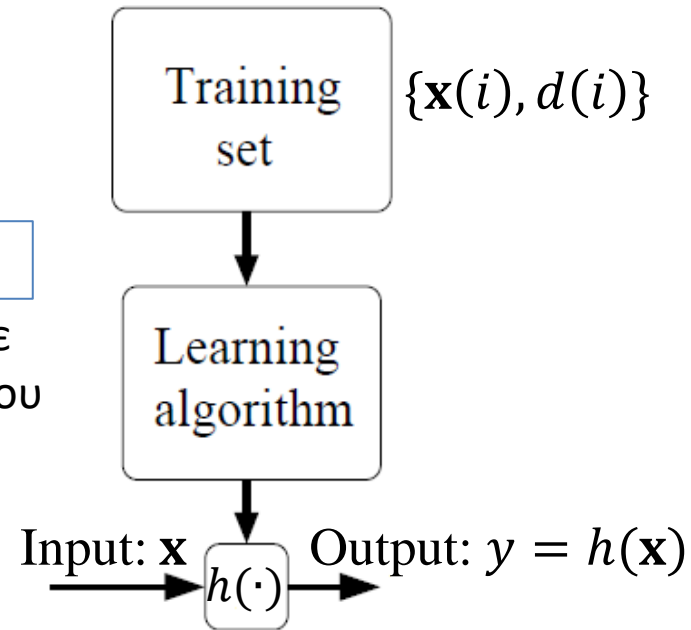
Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning (Επανάληψη)

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 επιθυμητή (**desired**) έξοδο d συμβατή με δεδομένη υπόθεση (π.χ. πρόβλεψη ή ταξινόμηση). Τα στοιχεία x_i είναι αριθμητικές τιμές που κωδικοποιούν m ειδοποιά χαρακτηριστικά (**features**) του \mathbf{x}

Ζητείται ο προσδιορισμός της συνάρτησης $y = h(\mathbf{x}) \cong d$

- Η σχεδίαση της $h(\cdot)$ προκύπτει από αλγόριθμο μάθησης, με προσαρμογή της μορφής και των παραμέτρων ενός μοντέλου σε εμπειρικά δεδομένα ενός συνόλου N χαρακτηρισμένων (**labeled**) του δείγματος μάθησης (**Training Set**) ζευγών, $\{\mathbf{x}(i), d(i)\}$, $i = 1, 2, \dots, N$ ώστε να προσεγγίζεται ο στόχος της υπόθεσης $d(i) \cong y(i) = h(\mathbf{x}(i))$
- 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

Μη Επιβλεπόμενη Μάθηση – *Unsupervised Learning, Hebbian Learning*

- Εκτίμηση *a-priori* πιθανοτήτων $p(\mathbf{x})$ του δειγματικού δεδομένου $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ με m *features* x_i , π.χ. *K-Means Clustering* - επιλογή K Κέντρων Βάρους Συστοιχιών (*cluster centroids*) και κατανομή σε *clusters* δυσδιάστατων σημείων $\mathbf{x} = [x_1, x_2]^T$
- Δεν υπάρχει εκ των προτέρων κατηγοροποίηση (*labelling*) δεδομένων. Το σύστημα με βάση *features* δειγματικών στοιχείων \mathbf{x} (π.χ. συντεταγμένες σημείων σε δύο διαστάσεις) ορίζει επιθυμητή έξοδο d και χειρίζεται νέα στοιχεία ως προς αυτή (π.χ. συμμετοχή στο πλησιέστερο *cluster centroid*) μέσω μετασχηματισμού (συνάρτησης) $y = h_{\mathbf{w}}(\mathbf{x}) \rightarrow d$ με στατιστικά κριτήρια
- Τα συστήματα με μη επιβλεπόμενη μάθηση βασίζονται σε αυτο-οργάνωση (π.χ. *Self-Organizing Maps - SOM*) και μαθηματικά εργαλεία ανίχνευσης κοινών χαρακτηριστικών (*features, patterns*) για την αποτελεσματική επεξεργασία – αποθήκευση – ταξινόμηση δεδομένων χωρίς εξωτερική βοήθεια, π.χ. για επεξεργασία σήματος (*speech – image processing*) και αναγνώριση προτύπων (*pattern recognition*)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διαμόρφωση Συστάδων μέσω K - Means Clustering

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

Ορισμοί:

- **Encoder** $C(i) = j$: Το στοιχείο $\mathbf{x}(i)$, $i = 1, 2, \dots, N$ ανήκει στο *cluster* $j = 1, 2, \dots, K$
- **Measure of Similarity** $d(\mathbf{x}(i), \mathbf{x}(i'))$ π.χ. **Ευκλείδεια απόσταση** $\|\mathbf{x}(i) - \mathbf{x}(i')\|^2$
- **Εκτιμώμενο κέντρο βάρους** $\hat{\mu}_j$ (*centroid*) του *cluster* $j = 1, 2, \dots, K$
- **Συνάρτηση Κόστους**

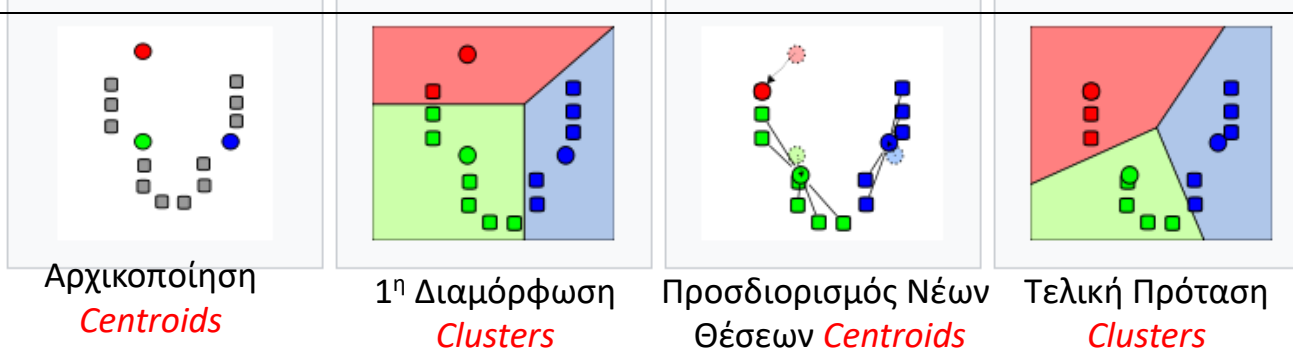
$$J(C) = \frac{1}{2} \sum_{j=1}^K \sum_{C(i)=j} \sum_{C(i')=j} \|\mathbf{x}(i) - \mathbf{x}(i')\|^2 = \sum_{j=1}^K \sum_{C(i)=j} \|\mathbf{x}(i) - \hat{\mu}_j\|^2$$

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

Σύνοψη αλγορίθμου:

- Ο αλγόριθμος τοποθετεί τα δειγματικά στοιχεία στο πλησιέστερο *centroid*, ανανεώνει τις επιλογές κέντρων βάρους (*centroids*) και συνεχίζει μέχρι τη τελική διαμόρφωση K *clusters*
- Αρχική επιλογή *centroids*: Αυθαίρετες (τυχαίες) τοποθετήσεις K σημείων
- Η εκτέλεση του αλγορίθμου είναι αποτελεσματική αλλά χωρίς εγγύηση βέλτιστης λύσης

Παράδειγμα: $K = 3, N = 12$ (https://en.wikipedia.org/wiki/K-means_clustering)



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Μη Επιβλεπόμενη Μάθηση: Ανάλυση Κυρίων Συνιστωσών

Principal Components Analysis - PCA

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

Αναλογία με ιδιοτιμές (*eigenvalues*) και ιδιοδιανύσματα (*eigenvectors*) στη γραμμική άλγεβρα και σχετικούς αλγόριθμους μετασχηματισμού συνιστωσών σε ορθοκανονικό (*orthonormal*) σύστημα συντεταγμένων

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Orthonormal Transformation σε Principal Components

Η υπόλοιπη ύλη βασίζεται στο Simon Haykin, "Neural Networks & Learning Machines", Pearson 2009

Ορισμοί:

- Δειγματικό στοιχείο δεδομένων \mathbf{x} με m χαρακτηριστικά (*features*) που κωδικοποιούνται στις συνιστώσες x_i : $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$
- Το στοιχείο \mathbf{x} ανήκει στον δειγματικό χώρο *τυχαίων* διανυσμάτων \mathbf{X} και x_i είναι δειγματικές τιμές της τυχαίας μεταβλητής X_i . Θεωρούμε 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 λ_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_i τυχαία μεταβλητή με δειγματική τιμή a_j
- Οι αρχικές συντεταγμένες προκύπτουν μονοσήμαντα από τις Κύριες Συνιστώσες:

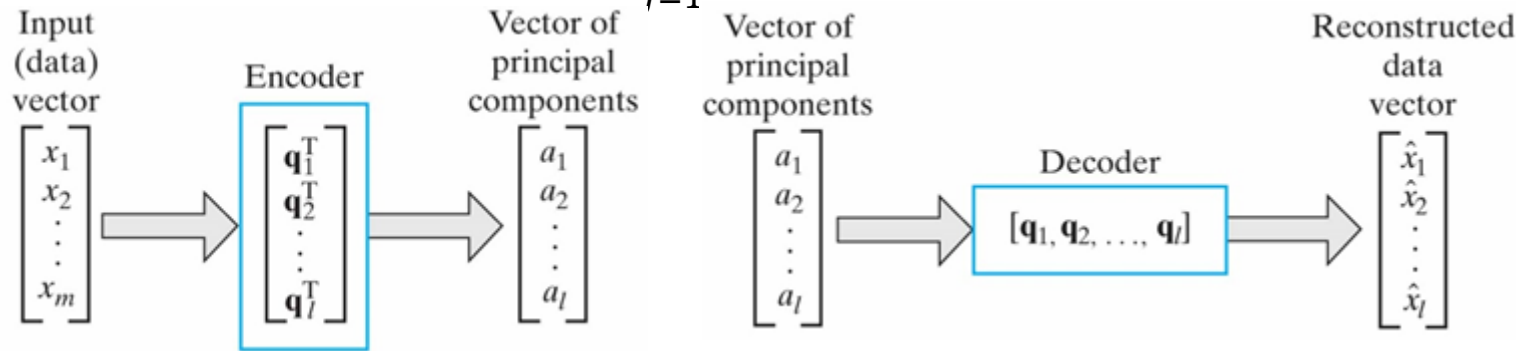
$$\mathbf{x} = \sum_{j=1}^m a_j \mathbf{q}_j$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Orthonormal Transformation σε Principal Components

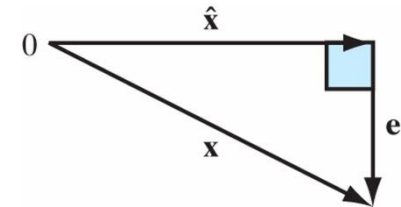
Αν αγνοήσουμε τις συνιστώσες $\alpha_j, j = l + 1, l + 2, \dots, m$ προκύπτει εκτίμηση του στοιχείου \mathbf{x} από το στοιχείο $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_l] = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_l][a_1, a_2, \dots, a_l]^T$ με μικρότερο αριθμό συνιστωσών $l < m$ επιλέγοντας τις σημαντικότερες συνιστώσες - *Principal Components*

$$\hat{\mathbf{x}} = \sum_{i=1}^l a_i \mathbf{q}_i \text{ για } l < m$$



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 τυχαίων μεταβλητών X_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

Προσδιορισμός Μέγιστης Συνιστώσας μέσω Φίλτρου Hebbian ενός Γραμμικού Νευρώνα (1/2) (First Principal Component via Hebbian-based Maximum Eigenfilter)

Αξίωμα του Hebb's: Αν τα σήματα στα άκρα μιας νευρωνικής σύναψης i ενεργοποιούνται ταυτόχρονα σε μια χρονική στιγμή 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 Components** από Δεδομένα m Διαστάσεων (Features)

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

Τα βάρη αυξάνονται στην επανάληψη $n \rightarrow n + 1$ αν το γινόμενο $y(n)x_i(n)$ είναι θετικό (*αξίωμα Hebb's*)

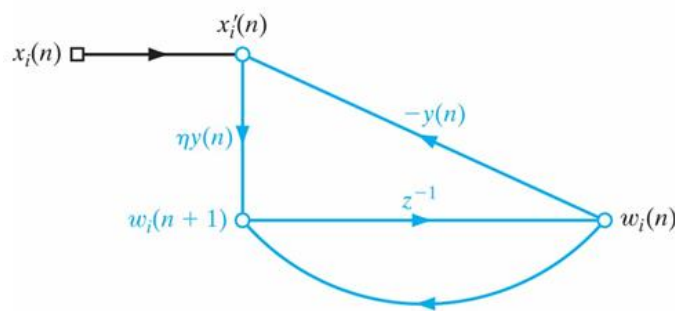
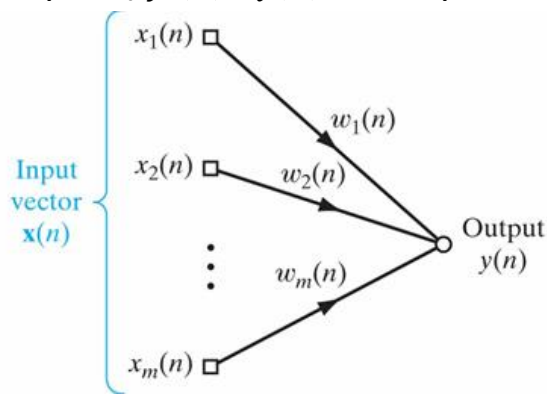
$$w_i(n + 1) = w_i(n) + \eta y(n)x_i(n), \quad i = 1, 2, \dots, m \text{ όπου } \eta \text{ learning parameter}$$

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

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

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

$$x'_i(n) = x_i(n) - y(n)w_i(n), \quad i = 1, 2, \dots, m$$



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Προσδιορισμός Μέγιστης Συνιστώσας μέσω Φίλτρου Hebbian ενός Γραμμικού Νευρώνα (2/2) (First Principal Component via Hebbian-based Maximum Eigenfilter)

Ζητήματα Σύγκλισης

Διανυσματικοί ορισμοί

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

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

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

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 λόγω της μη επιβλεπόμενης (αυτό-οργανωμένης) φύσης του πλην του a-priori καθορισμού της παραμέτρου μάθησης η

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

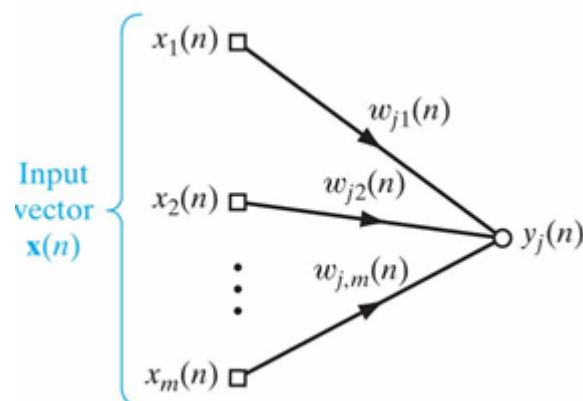
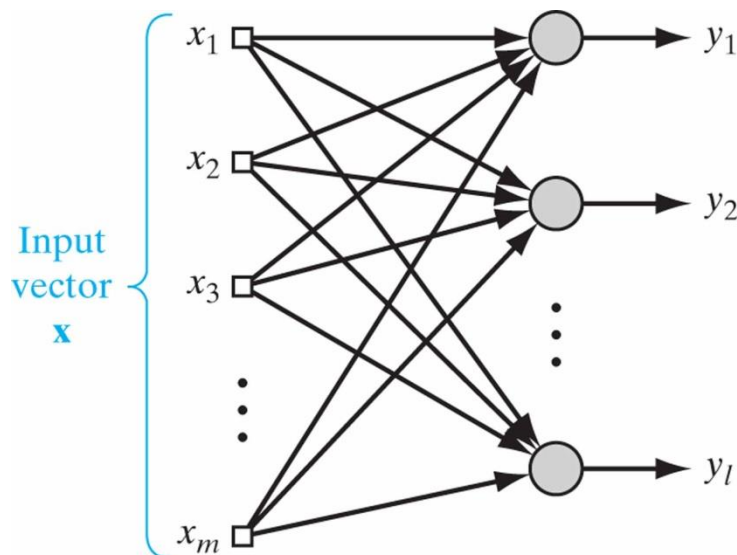
Ανάλυση Κύριων Συνιστωσών μέσω Γραμμικού Μονοεπίπεδου Νευρωνικού Δικτύου (1/3) (Hebbian-based Principal-Component Analysis)

Γενίκευση του Hebbian-based maximum Eigenfilter: Γραμμικό Feedforward Νευρωνικό Δίκτυο ενός επιπέδου με l νευρώνες για προσδιορισμό των l 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 στοιχείο εισόδου $x_i(n)$, $i = 1, 2, \dots, m$ προς την j κύρια συνιστώσα $y_j(n)$, $j = 1, 2, \dots, l$ διαφοροποιούνται στην επανάληψη $n \rightarrow n + 1$ κατά $\Delta w_{ji}(n)$:

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



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ανάλυση Κύριων Συνιστωσών μέσω Γραμμικού Μονοεπίπεδου Νευρωνικού Δικτύου (2/3)

(Hebbian-based Principal-Component Analysis)

Γενικευμένος Αλγόριθμος του Hebb's – Generalized Hebbian Algorithm (GHA)

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

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

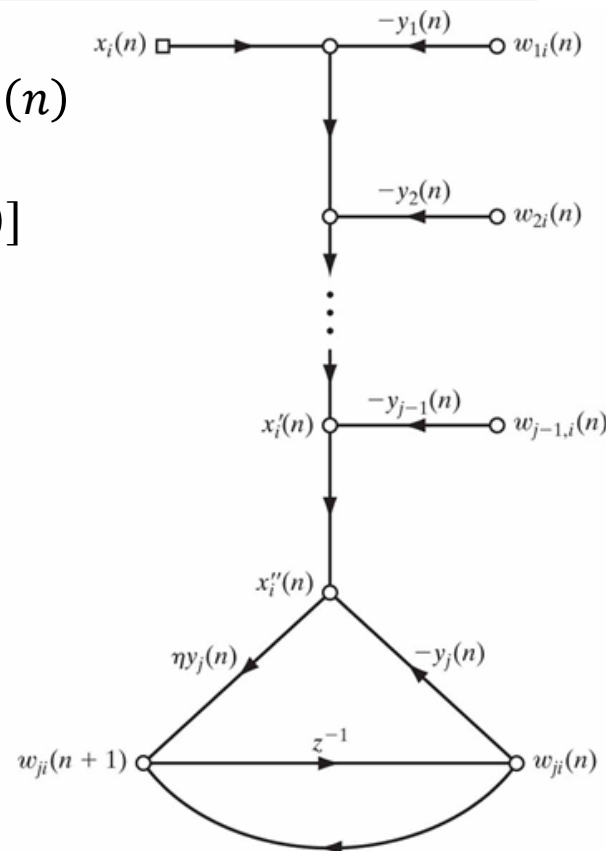
$$\Delta w_{ji}(n) = \eta y_j(n) x''_i(n) \text{ όπου } x''_i(n) = x'_i(n) - w_{ji}(n) y_j(n)$$

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

Διανυσματική Μορφή:

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

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



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ανάλυση Κύριων Συνιστωσών μέσω Γραμμικού Μονοεπίπεδου Νευρωνικού Δικτύου (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 $y_1(n)$

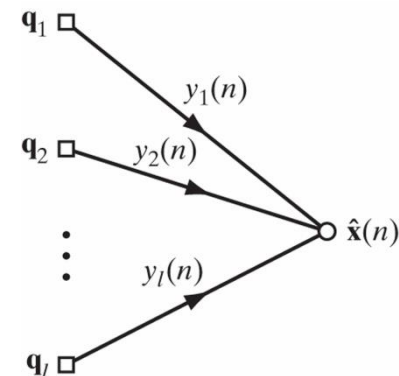
Για $j = 3$: $\mathbf{x}'(n) = \mathbf{x}(n) - \mathbf{w}_1(n) y_1(n) - \mathbf{w}_2(n) y_2(n)$

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 $\mathbf{R} = \mathbf{E}[\mathbf{X}\mathbf{X}^T]$, $k = 1, 2, \dots, l$ 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 δειγματικό στοιχείο αναλύεται σε γραμμικό συνδυασμό $m = 64$ features, που ορίζουν διανύσματα εισόδου κωδικοποιημένα σε 256 gray levels:

$$\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_m(n)]^T \quad n = 1, 2, \dots, 2000$$

- Το δείγμα τροφοδοτεί **Μονοστρωματικό Linear Feedforward Δίκτυο** με $l = 8$ Νευρώνες εξόδου
- Το σύστημα συγκλίνει σε 8×64 συναπτικά βάρη $w_{ji}(n)$ και απολήγει σε 8 εξόδους (κυριότερες συνιστώσες, **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 $8 \times 64 = 1024$ 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 (**Principal Components**) χωρίς κβαντισμό

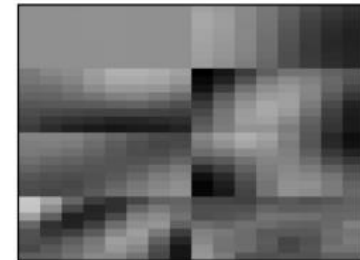
$$\hat{\mathbf{x}}(n) = \sum_{k=1}^l y_k(n) \mathbf{q}_k, \quad \mathbf{q}_k = \lim_n \mathbf{w}_k(n), \quad \mathbf{w}_k(n) = [w_{k1}(n), w_{k2}(n), \dots, w_{km}(n)]^T$$

- 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



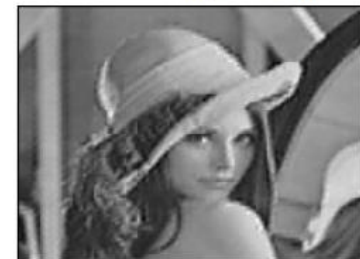
Weights



Using First 8 Components



11 to 1 compression



ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

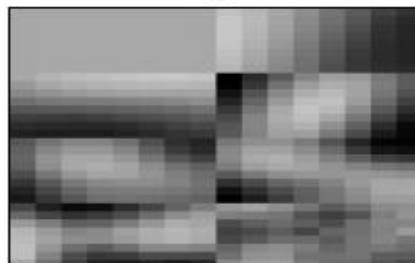
Εφαρμογή PCA μέσω Generalized Hebbian Algorithm (GHA) στη Κωδικοποίηση Εικόνας (2/2)

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

Original Image



Weights



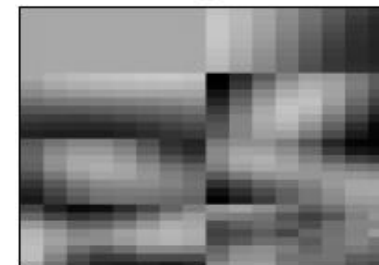
Using First 8 Components



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



Weights



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



Weights

