6. Ανάλυση χαρακτηριστικών

Μια σημαντική εργασία σε ένα σύστημα αναγνώρισης είναι η ανάλυση των μετρούμενων χαρακτηριστικών των προτύπων. Με την ανάλυση των χαρακτηριστικών πετυχαίνουμε την αξιολόγηση τους, την επιλογή των καταλληλότερων για την ταξινόμηση των προτύπων και την ελάττωση του πλήθους τους. Για παράδειγμα ένα χαρακτηριστικό με κοντινές τιμές μέσα σε κάθε μία κλάση και σαφώς διαφορετικές τιμές μεταξύ των κλάσεων είναι «καλό» χαρακτηριστικό ταξινόμησης. Με κατάλληλες μεθόδους ανάλυσης χαρακτηριστικών μπορούμε να οδηγηθούμε και σε νέα χαρακτηριστικά που προκύπτουν από τον γραμμικό συνδυασμό των αρχικών και είναι καταλληλότερα από αυτά. Ακολούθως θα δούμε μεθόδους ανάλυσης χαρακτηριστικών σε συστήματα εκμάθησης με και χωρίς επόπτη.

6.1. Ανάλυση χαρακτηριστικών στην εκπαίδευση με επόπτη

Μια σημαντική εργασία σε ένα σύστημα αναγνώρισης είναι η επιλογή ενός μικρού συνόλου κατάλληλων χαρακτηριστικών από ένα πολύ μεγαλύτερο. Η επιλογή των ισχυρών χαρακτηριστικών είναι κρίσιμη για την αποτελεσματικότητα της ταξινόμησης. Για να ελαττωθεί το πλήθος των χαρακτηριστικών με την επιλογή ενός συνόλου «καλών», που οδηγούν στην ορθή εκτελείται μια διαδικασία αξιολόγησης της *ευστάθειας (stability)*, της *διαχωριστικότητας (separability)* και της *ομοιότητας (similarity)* των χαρακτηριστικών.

Με τον έλεγχο της ευστάθειας προσδιορίζονται τα χαρακτηριστικά που παρουσιάζουν σημαντικά σταθερή συμπεριφορά. Υπολογίζονται οι μέσες τιμές και οι τυπικές αποκλίσεις των χαρακτηριστικών μέσα σε κάθε κλάση. Οι τιμές των τυπικών αποκλίσεων κανονικοποιούνται κατόπιν στο ίδιο διάστημα, διαιρούμενες με τις μέσες τιμές τους. Τα χαρακτηριστικά διαιρούνται σε τρεις ομάδες ανάλογα με την κανονικοποιημένη τυπική απόκλισή τους σ^*. Η πρώτη ομάδα περιλαμβάνει τα χαρακτηριστικά υψηλής ευστάθειας με $\sigma^*<0.1$. Η δεύτερη ομάδα περιλαμβάνει τα ασταθή χαρακτηριστικά με $0.1<\sigma^*<0.9$ και η τρίτη τα πολύ ασταθή με $\sigma^*>0.9$.

Για τον έλεγχο της ικανότητας διαχωρισμού καθορίζεται ο *παράγοντας διαχωριστικότητας (seperability factor)* S_{avg}^{ℓ} μεταξύ δύο κλάσεων ω_i και ω_j για κάθε χαρακτηριστικό *ν* από τη σχέση:

$$S_\nu = \frac{\left|\mu_\nu^i - \mu_\nu^j\right|}{\sqrt{(\sigma_\nu^i)^2 + (\sigma_\nu^j)^2}} \tag{6.1.1}$$

όπου μ_ν^i, μ_ν^j είναι οι μέσες τιμές και σ_ν^i, σ_ν^j οι τυπικές αποκλίσεις σε κάθε κλάση. Μεγάλη διαχωριστικότητα σημαίνει ότι το χαρακτηριστικό έχει μεγάλη ικανότητα να διαχωρίζει τις δύο κλάσεις μεταξύ τους.

Για την ανάλυση της ομοιότητας των χαρακτηριστικών εκτιμάται ο *παράγοντας συσχέτισης (correlation factor)* για κάθε δύο χαρακτηριστικά *ν* και *λ* που ανήκουν στην ίδια τάξη *p*, σύμφωνα με τη σχέση

$$C_{\nu\lambda}^p = \frac{\frac{1}{K_p}\sum_{\kappa=1}^{K_p}(x_{\nu\kappa} - \mu_\nu)(x_{\lambda\kappa} - \mu_\lambda)}{\sigma_\nu \sigma_\lambda} \tag{6.1.2}$$

όπου K_p είναι το πλήθος των στοιχείων της κλάσης *p*, $x_{\nu\kappa}$ και $x_{\lambda\kappa}$ οι τιμές των χαρακτηριστικών, $\mu_\nu, \mu_\lambda, \sigma_\nu$ και σ_λ οι μέσες τιμές και οι τυπικές αποκλίσεις των χαρακτηριστικών στην τάξη *p*. Ο παράγοντας συσχέτισης μετρά την ομοιότητα μεταξύ των δύο χαρακτηριστικών και παίρνει τιμές μεταξύ *–1* και *+1*. Μια τιμή κοντά στο *+1* ή στο *-1* σημαίνει ότι τα δύο χαρακτηριστικά συσχετίζονται πολύ ή συσχετίζονται αντίστροφα, αντίστοιχα. Μια τιμή κοντά στο μηδέν δείχνει ότι τα χαρακτηριστικά είναι κατά πολύ ασυσχέτιστα.

6.2. Ανάλυση χαρακτηριστικών στην εκπαίδευση χωρίς επόπτη

Η ανάλυση χαρακτηριστικών στην εκπαίδευση χωρίς επόπτη, πετυχαίνεται με την μέθοδο *ανάλυσης κύριων συνιστωσών* (*ΑΚΣ*), (*PCA*: *Principal Components Analysis).* Η ΑΚΣ είναι ένα μέσο με το οποίο επιτυγχάνεται ένας κατάλληλος μετασχηματισμός ώστε ο χώρος των χαρακτηριστικών να εμφανίζει τη μεγαλύτερη δυνατή διασπορά. Ειδικά με τη χρήση του ΑΚΣ ο αρχικός χώρος των χαρακτηριστικών

μετασχηματίζεται σε έναν άλλο χώρο που έχει την ίδια διάσταση με τον αρχικό. Ωστόσο ο μετασχηματισμός έχει σχεδιαστεί ώστε να είναι δυνατή η αντιπροσώπευση του αρχικού συνόλου των χαρακτηριστικών με άλλα λιγότερα και αποτελεσματικότερα, που διατηρούν την περισσότερη πληροφορία των δεδομένων. Συνεπώς, με τη χρήση της ΑΚΣ επιτυγχάνεται όχι μόνο η αύξηση της διασποράς των χαρακτηριστικών αλλά και η μείωση των διαστάσεων. Ο μετασχηματισμός Karhunen-Loeve (**KLT**) είναι ένα γνωστό εργαλείο στην ανάλυση πολλών μεταβλητών και χρησιμοποιείται στην ΑΚΣ.

Για ένα σύνολο *K* ανυσμάτων σε *ν*-διαστάσεων χώρο χαρακτηριστικών έστω ότι $\boldsymbol{x}_\kappa$ είναι το διάνυσμα που δείχνεται από την σχέση.

(6.2.1) $$\mathbf{x}_\kappa = \begin{bmatrix} x_{1\kappa} \\ x_{2\kappa} \\ \vdots \\ x_{\nu\kappa} \\ \vdots \\ x_{N\kappa} \end{bmatrix}$$

και $\boldsymbol{\mu}_x$ το άνυσμα μέσης τιμής των στοιχείων του συνόλου εκπαίδευσης

(6.2.2) $$\mu_x = E[\mathbf{x}_\kappa] = \frac{1}{K}\sum_{\kappa=1}^{K}\mathbf{x}_\kappa$$

όπου *E[.]* η μαθηματική προσδοκία. Τις τιμές x_κ-μ_x συμβολίζουμε με την διανυσματική μεταβλητή $x=x_\kappa-\mu_x$. Ο πίνακας συνδιασποράς $\boldsymbol{C}_x$ όλων των ανυσμάτων του συνόλου εκπαίδευσης δίνεται από τη σχέση

(6.2.3) $$\mathbf{C}_x = E[\mathbf{x}\mathbf{x}^T] = \frac{1}{K}\sum_{\kappa=1}^{K}(\mathbf{x}_\kappa - \mu_x)(\mathbf{x}_\kappa - \mu_x)^T =$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & . & \sigma_{1\nu} & . & . & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & . & . & \sigma_{2\nu} & . & . & \sigma_{2N} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ \sigma_{\nu 1} & \sigma_{\nu 2} & . & . & \sigma_\nu^2 & . & . & \sigma_{\nu N} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ \sigma_{N1} & \sigma_{N2} & . & . & \sigma_{N\nu} & . & . & \sigma_N^2 \end{bmatrix}$$

Κάθε διαγώνια τιμή σ_ν^2 του πίνακα συνδιασποράς εκφράζει τη διασπορά των ανυσμάτων $\boldsymbol{x}_\kappa$ για τη μεταβλητή-άξονα ν, ενώ οι υπόλοιπες τη συνδιασπορά των δεδομένων μεταξύ δύο διαφορετικών μεταβλητών-αξόνων. Είναι φανερό πως για τη διαδικασία της ταξινόμησης είναι επιθυμητές μεγάλες τιμές των διαγωνίων τιμών σ_ν^2 διότι μαρτυρούν ένα μεγάλο άπλωμα (spreading) των δεδομένων κατά μήκος του κ άξονα, ενώ είναι επιθυμητές μηδενικές τιμές για τις συνδιασπορές ώστε οι μεταβλητές - άξονες να είναι μεταξύ τους ασυσχέτιστες. Για να επιτύχουμε αυτό αναζητούμε ένα τέτοιο μετασχηματισμό $\boldsymbol{W}$ που θα εκφράζεται από ένα πίνακα pxp ώστε τα διανύσματα $\mathbf{y} = \mathbf{W} \cdot \mathbf{x}$ να έχουν πίνακα συνδιασποράς $\boldsymbol{C_y}$ με την ακόλουθη μορφή:

(6.2.4) $$\mathbf{C_y} = \begin{bmatrix} \lambda_1 & 0 & . & . & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & \lambda_\nu & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & 0 & . & . & \lambda_N \end{bmatrix} = \boldsymbol{\Lambda}, \ \lambda_1 > \lambda_2 > \ldots > \lambda_\nu > \ldots > \lambda_N$$

Ακόμη ισχύει ότι

(6.2.5) $$\mathbf{C_y} = E[yy^T] = E[\mathbf{W}^T\mathbf{x}(\mathbf{W}^T\mathbf{x})^T] = E[\mathbf{W}T\mathbf{x}\mathbf{x}^T\mathbf{W}] = \mathbf{W}^T E[\mathbf{x}\mathbf{x}^T]\mathbf{W} = \mathbf{W}^T\mathbf{C}_x\mathbf{W}$$

Είναι φανερό ότι η τιμή $\lambda_\nu = \sigma_\nu'^2$ είναι η διασπορά του ***y*** κατά τον *ν* άξονα, ο οποίος εκφράζεται από την *ν* στήλη του πίνακα ***W*.** Η γραμμή $\mathbf{w}_\nu$ είναι "αποτελεσματικό" χαρακτηριστικό ενώ η τιμή λ_ν αξιολογεί τη σπουδαιότητά του. Ο πίνακας $\mathbf{C}_x$ είναι συμμετρικός και ο πίνακας **Λ** διαγώνιος. Η σχέση $\mathbf{W}^T\mathbf{C}_x\mathbf{W} = \mathbf{\Lambda}$ ικανοποιείται από τα ιδιοανύσματα $\mathbf{w}_1,\ldots,\mathbf{w}_N$ και τις ιδιοτιμές $\lambda_1,\ldots,\lambda_N$ του πίνακα $\mathbf{C}_\mathbf{x}$ με $\mathbf{W}=[\ \mathbf{w}_1,\ldots,\mathbf{w}_N]$ (Παράρτημα Β). Επειδή ο πίνακας $\mathbf{C}_x$ είναι συμμετρικός τα ιδιοδιανύσματα είναι κάθετα μεταξύ τους. **Οι** ιδιοτιμές είναι ρίζες του χαρακτηριστικού πολυωνύμου όπως αυτό προκύπτει από την σχέση

(6.2.6) $\mathrm{Det}(\mathbf{C}_x-\lambda\mathbf{I})=\mathbf{0}$

Για κάθε ιδιοτιμή λ_ν το αντίστοιχο ιδιοδιάνυσμα $\mathbf{w}_\nu$ είναι μία από τις άπειρες λύσεις του αόριστου συστήματος $(\mathbf{C}_x-\lambda_\nu\mathbf{I})\mathbf{w}_\nu=\mathbf{0}$. Είναι σκόπιμο να επιλέγουμε τα μοναδιαία ιδιοδιανύσματα ($\|\mathbf{w}_\nu\| = 1$). Σε περίπτωση που απαιτείται επιπλέον μείωση του πλήθους N των ιδιοδιανυσμάτων μπορούμε να επιλέξουμε τα M από αυτά (M<N) με τις μεγαλύτερες αντίστοιχες ιδιοτιμές, αποδεχόμενοι ένα μέσο τετραγωνικό σφάλμα

(6.2.7) $$\overline{\varepsilon^2}(m) = \sum_{\nu=M+1}^{N}\lambda_\nu\ .$$

<u>ΠΑΡΑΔΕΙΓΜΑ</u>

$$x_1 = \begin{bmatrix}1\\0\end{bmatrix},\ x_2 = \begin{bmatrix}1\\1\end{bmatrix},\ x_3 = \begin{bmatrix}2\\0\end{bmatrix},\ x_4 = \begin{bmatrix}4\\3\end{bmatrix},\ x_5 = \begin{bmatrix}5\\2\end{bmatrix}$$

$$\mu = \begin{bmatrix}(1+1+2+4+5)/5\\(0+1+0+3+4)/5\end{bmatrix} = \begin{bmatrix}\frac{13}{5}\\\frac{6}{5}\end{bmatrix} = \begin{bmatrix}2{,}6\\1{,}2\end{bmatrix}$$

$$C_x = \begin{vmatrix}\sigma_1^2 & \sigma_{12}^2\\\sigma_{21}^2 & \sigma_{22}^2\end{vmatrix}$$

$$\sigma_1^2 = \frac{(1-2.6)^2+(1-2.6)^2+(2-2.6)^2+(4-2.6)^2+(5-2.6)^2}{5} = 2.64$$

$$\sigma_{21}^2 = \frac{(0-1.2)(1-2.6)+(1-1.2)(1-2.6)+(0-1.2)(2-2.6)+(3-1.2)(4-2.6)+(2-1.2)(5-2.6)}{5} = 1.48$$

$$\sigma_2^2 = \frac{(0-1.2)^2 + (1-1.2)^2 + (0-1.2)^2 + (3-1.2)^2 + (2-1.2)^2}{5} = 1.36$$

$$\sigma_{12}^2 = \sigma_{21}^2 = 1.48$$

$$C_x = \begin{bmatrix} 2.64 & 1.48 \\ 1.48 & 1.36 \end{bmatrix}$$

$$\det(C_x - \lambda \cdot I) = \det\begin{bmatrix} 2.64-\lambda & 1.48 \\ 1.48 & 1.36-\lambda \end{bmatrix} = (2.64-\lambda)\cdot(1.36-\lambda) - 1.48^2 =$$

$$1.36 \cdot 2.64 - 1.36 \cdot \lambda - 2.64 \cdot \lambda + \lambda^2 - 1.48^2 =$$

$\lambda^2 - 4 \cdot \lambda - 1.4$ που είναι το χαρακτηριστικό πολυώνυμο.

Για $\lambda^2 - 4 \cdot \lambda - 1.4 = 0$ προκύπτουν οι ρίζες $\lambda_1 = 3.61$, $\lambda_2 = 0.39$ που είναι οι ιδιοτιμές του C_x. Τα αντίστοιχα ιδιοδυανύσματα $\mathbf{w}_1$, $\mathbf{w}_2$ ικανοποιούν τις σχέσεις

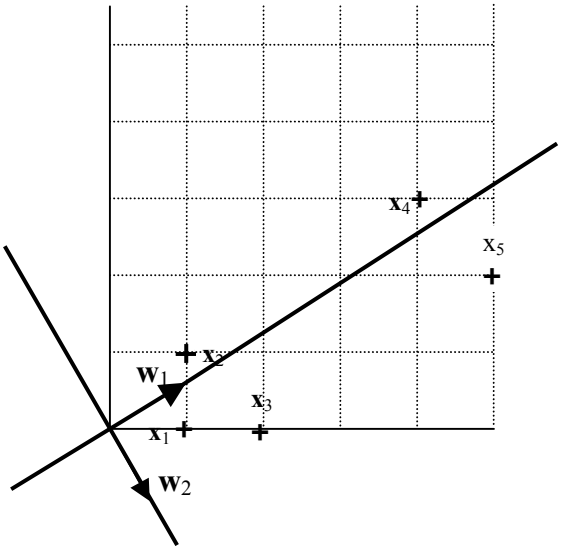
(6.2.8) $(\mathbf{C}_x - \lambda_1 \cdot \mathbf{I}) \cdot \mathbf{w}_1 = 0$ και

(6.2.9) $(\mathbf{C}_x - \lambda_2 \cdot \mathbf{I}) \cdot \mathbf{w}_2 = 0$.

Για το $\mathbf{w}_1$ θα έχουμε αναλυτικά

$$\begin{vmatrix} 2.64-3.61 & 1.48 \\ 1.48 & 1.36-3.61 \end{vmatrix} \cdot \begin{vmatrix} w_{11} \\ w_{21} \end{vmatrix} = 0 \Rightarrow \begin{matrix} -0.97 \cdot w_{11} + 1.48 \cdot w_{21} = 0 \\ 1.48 \cdot w_{11} - 2.25 \cdot w_{21} = 0 \end{matrix}$$

από το παραπάνω αόριστο σύστημα λύση για $w_{11}=\alpha$ είναι $\mathbf{w}_1=[\alpha, \alpha \cdot 0.65]^T$. Ακόμη για να ισχύει $\|\mathbf{w}_1\| = 1 \Leftrightarrow \sqrt{\alpha^2 + (\alpha \cdot 1.65)^2} = 1 \Leftrightarrow \alpha = \pm 1/\sqrt{1+1.65^2} \approx .84$, άρα ένα μοναδιαίο ιδιοδιάνυσμα είναι το $\mathbf{w}_1=[0.84,0.55]^T$. Όμοια για την ιδιοτιμή λ_2 προκύπτει $\mathbf{w}_2=[0.55,-0.84]^T$. Τα διανύσματα είναι κάθετα μεταξύ τους όπως αναμενόταν. Στο Σχ.6.2-1 φαίνεται η ορθότητα της μεθόδου.



Σχήμα 6.2-1

Ο KLT είναι μία αναλυτική διαδικασία για την εκτίμηση των ιδιοτιμών και των ιδιοδιανυσμάτων από τον πίνακα συνδιασποράς των δεδομένων $\boldsymbol{C_x}$. Ειδικά, εάν το πλήθος των διαστάσεων του χώρου των ανυσμάτων είναι μεγάλο ο υπολογισμός και ο χειρισμός του πίνακα $\boldsymbol{C_x}$ είναι ανέφικτος.

Για τον παραπάνω λόγο αντί του KLT μπορούμε να χρησιμοποιήσουμε ένα γραμμικό νευρωνικό δίκτυο ενός επιπέδου (single-layer linear feed forward neural net), το οποίο επιτελεί την ΑΚΣ. Το δίκτυο αυτό εκπαιδεύεται χωρίς επόπτη (unsupervised) από τον γενικευμένο αλγόριθμο του Hebb που βασίζεται στον κανόνα εκμάθησης του Hebb. Με τον αλγόριθμο αυτό το νευρωνικό δίκτυο συγκλίνει με πιθανότητα ένα στα ιδιοδιανύσματα ενός πίνακα συνδιασποράς $\boldsymbol{C_x}$ απεριόριστου μεγέθους. Ο υπολογισμός του $\boldsymbol{C_x}$ δεν είναι αναγκαίος επειδή τα ιδιοδιανύσματα προκύπτουν κατευθείαν από τα δεδομένα. Η εκπαίδευση του νευρωνικού δικτύου γίνεται ως εξής:

Έστω ότι

- $t=1,2,3,\ldots$ μία μεταβλητή για τη μέτρηση της επανάληψης της διαδικασίας,
- $\boldsymbol{x}(t)$ το διάνυσμα εισόδου στο νευρωνικό δίκτυο τη χρονική στιγμή t, με συνιστώσες $x_i(t)$, $i=0,1,\ldots,p-1$,
- m το πλήθος των νευρώνων ($m \le p$) που ως εκ τούτου είναι και το πλήθος των επιθυμητών κύριων συνιστωσών,
- j ένας δείκτης που αποδίδεται στους νευρώνες $j \in \{0,1,\ldots,m-1\}$,
- $w_{ji}(t)$ η τιμή του βάρους της σύναψης που συνδέει τον j νευρώνα με την i είσοδο κατά την επανάληψη t,
- $\boldsymbol{W_{m\,x\,p}}(t)$ ο πίνακας των $w_{ji}(t)$ και
- $\boldsymbol{y}(t)$ το άνυσμα εξόδου του δικτύου κατά την επανάληψη t, με συνιστώσες $y_j(t)$.

Εκτελούνται τα ακόλουθα βήματα υπολογισμών:

Βήμα 1. Αφαιρείται πρώτα το διάνυσμα $\boldsymbol{\mu_x}$ της μέσης τιμής από κάθε στοιχείο του συνόλου εκπαίδευσης. Με τον τρόπο αυτό οι τιμές που προκύπτουν έχουν μηδενικό διάνυσμα μέσης τιμής.

Βήμα 2. Αποδίδονται αρχικά ($t=0$) στα βάρη $w_{ji}(0)$ των συνάψεων μικρές τυχαίες τιμές και στην παράμετρο γ του ρυθμού εκπαίδευσης μικρή θετική τιμή (π.χ. $\gamma=0.007$).

Βήμα 3. Υπολογίζεται το διάνυσμα εξόδου $\boldsymbol{y}(t)$ με συνιστώσες $y_j(t)$ για $i=0,1,...,p-1$, $j=0,1,...,m-1$ από τη σχέση

(6.2.8) $$y_j(t)=\sum_{i=0}^{p-1} w_{ji}(t)\cdot x_i(t), \text{ ή } \boldsymbol{y}(t)=\boldsymbol{W}(t)\,\boldsymbol{x}(t)$$

και η ποσότητα $\Delta w_{ji}(t)$ από την σχέση

(6.2.9) $$\Delta w_{ji}(t)=\gamma\left(\underbrace{y_j(t)\cdot x_i(t)}_{(\alpha)}-\underbrace{y_j(t)\cdot\sum_{k=0}^{j} w_{ki}(t)\cdot y_k(t)}_{(\beta)}\right)$$

Ο όρος (α) εκφράζει τον απλό κανόνα του Hebb, που λέει ότι εάν δύο νευρώνες που βρίσκονται στα άκρα μιας σύναψης (σύνδεσης), ενεργοποιούνται συγχρόνως, τότε η ισχύς της σύναψης αυξάνεται. Αλλιώς εξασθενεί ή εκφυλίζεται. Ο όρος (β) επιβάλλει ένα όριο στην αύξηση της σύναψης. Η τιμή του $w_{ji}(t+1)$ προσαρμόζεται σύμφωνα με τη σχέση

(6.2.10) $$w_{ji}(t+1)= w_{ji}(t)+\Delta w_{ji}\,.$$

Βήμα 4. Αύξηση της μεταβλητής t κατά ένα και επανάληψη από το *βήμα 3* έως ότου τα βάρη των συνάψεων w_{ji} φθάσουν στη σταθερή κατάσταση.

Μετά τη φάση της εκπαίδευσης τα βάρη w_{ji} του j συγκλίνουν στην i συνιστώσα του ιδιοδιανύσματος που αντιστοιχεί στην j ιδιοτιμή του πίνακα συνδιασποράς $\boldsymbol{C_x}$. Με άλλα λόγια οι γραμμές του πίνακα $\boldsymbol{W}$ έχουν προσεγγίσει τα πρώτα m ιδιοδιανύσματα του $\boldsymbol{C_x}$, ταξινομημένες σε φθίνουσα σειρά.

ΠΑΡΑΡΤΗΜΑ Ι

Ο ανάστροφος πίνακα $\mathbf{A}^T$ ενός πίνακα **A** είναι ένας πίνακας που έχει ως γραμμές τις στήλες του **A** με ίδιο δείκτη. Αν ο $\mathbf{A}_{MxN}$ έχει M γραμμές και N στήλες, τότε ο $\mathbf{A}^T_{NxM}$ θα έχει N γραμμές και M στήλες. Η *ν* γραμμή του $\mathbf{A}^T$ είναι η *ν* στήλη του **A** και η *μ* γραμμή του η *μ* στήλη του **A.** Για παράδειγμα αν

$$\mathbf{A}_{3\times 2} = \begin{vmatrix} 1 & 0.5 \\ -1 & 3 \\ 2 & -2 \end{vmatrix} \Leftrightarrow \mathbf{A}^T_{2\times 3} = \begin{vmatrix} 1 & -1 & 2 \\ 0.5 & 3 & -2 \end{vmatrix}.$$

Πολλαπλασιασμός πινάκων. Έστω οι πίνακες $\mathbf{A}_{N\times K}$ και $\mathbf{B}_{K\times M}$, το γινόμενό τους είναι ένας πίνακας $\mathbf{\Gamma}_{N\times M}$ για τον οποίο γραφουμε $\mathbf{\Gamma}=\mathbf{A}\cdot\mathbf{B}$ και έχει στοιχεία $\Gamma_{\mu\kappa}$ που δίνονται από την σχέση

$$\Gamma_{\nu\mu} = \sum_{\kappa=1}^{K} A_{\nu\kappa} \cdot B_{\kappa\mu}$$

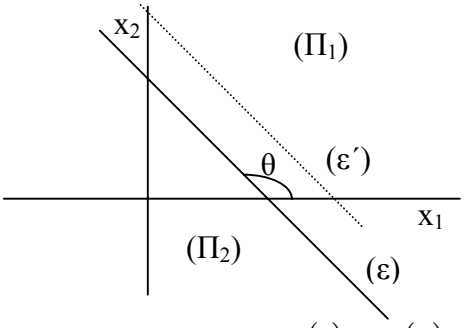
Για παράδειγμα αν

$$A_{3\times 2} = \begin{vmatrix} 1 & 2 \\ 0.5 & 4 \\ 3 & 1 \end{vmatrix} \text{ και } B_{2\times 3} = \begin{vmatrix} 3 & -1 & 1 \\ 1 & 0.5 & -0.5 \end{vmatrix} \text{ τότε}$$

$$A\cdot B = \Gamma_{3\times 3} = \begin{vmatrix} 1\cdot 3+2\cdot 1 & 1\cdot(-1)+2\cdot 0.5 & 1\cdot 1+2\cdot(-0.5) \\ 0.5\cdot 3+4\cdot 1 & 0.5\cdot(-1)+4\cdot 0.5 & 0.5\cdot 1+4\cdot(-0.5) \\ 3\cdot 3+1\cdot 1 & 3\cdot(-1)+1\cdot 0.5 & 3\cdot 1+1\cdot(-0.5)= \end{vmatrix} = \begin{vmatrix} 5 & 0 & 0 \\ 7.5 & 1.5 & -1.5 \\ 10 & -2.5 & 2.5 \end{vmatrix}$$

Εξισώσεις ευθείας και ανισώσεις ημιεπιπέδων στον E^2

Έστω η ευθεία (ε) του σχήματος με εξίσωση



$$(\varepsilon) \rightarrow d(x) = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 = 0$$

Η (ε) χωρίζει το καρτεσιανό επίπεδο σε ημιεπίπεδα. Η κλίση θ της (ε) έχει

$\varepsilon\varphi\theta = -\frac{w_1}{w_2}$. Η (ε) τέμνει τον άξονα x_2 στο σημείο $\left(0,-\frac{w_3}{w_2}\right)$.

Αν d(x)=d>0 $\Rightarrow$ d(x)-d=0 $\Rightarrow w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 - d = 0$

Η εξίσωση αυτή δίνει τα σημεία της ευθείας (ε΄) με κλίση θ΄ που έχει

$\varepsilon\varphi\theta' = -\frac{w_1}{w_2} = \varepsilon\varphi\theta$, άρα ε΄//ε. Η (ε') τέμνει τον άξονα των x_2 στο σημείο

$$\left(0,-\frac{w_3-d}{w_2}\right) = \left(0,-\frac{w_3}{w_2}+\frac{d}{w_2}\right)$$

Αν w_2>0 τότε όλα τα σημεία της (ε΄) βρίσκονται στο ημιεπίπεδο (Π_1) πάνω από το σημείο $\left(0,-\frac{w_3}{w_2}\right)$.

Για d $\in (0,+\infty)$ η (ε΄) σαρώνει όλο το ημιεπίπεδο (Π_1). Άρα η σχέση d(x)>0 επαληθεύεται από τα σημεία του ενός ημιεπιπέδου και η d(x)<0 του άλλου.

ΠΑΡΑΡΤΗΜΑ ΙΙ

Σε ένα συμμετρικό πίνακα $\mathbf{C}_{NXN}$, $\mathbf{C}^T = \mathbf{C}$, με ιδιοτιμές $\lambda_\mu, \lambda_\nu, \lambda_\mu \neq \lambda_\nu$ και αντίστοιχα μοναδιαία ιδιοδιανύσματα $\mathbf{w}_\mu \neq \mathbf{w}_\nu$ ισχύει ότι $\mathbf{w}_\mu^T \cdot \mathbf{w}_\nu = 0$. Δηλαδή τα ιδιοδιανύσματα είναι κάθετα μεταξύ τους.

Απόδειξη: Επειδή λ_μ, λ_ν ιδιοτιμές και $\mathbf{w}_\mu, \mathbf{w}_\nu$ ιδιοδιανύσματα ισχύουν οι σχέσεις

$$\left.\begin{array}{l}\mathbf{C}\cdot\mathbf{w}_\mu = \lambda_\mu \cdot \mathbf{w}_\mu \\ \mathbf{C}\cdot\mathbf{w}_\nu = \lambda_\nu \cdot \mathbf{w}_\nu\end{array}\right\} \Rightarrow \begin{array}{l}\mathbf{w}_\nu^T \cdot \mathbf{C}\cdot\mathbf{w}_\mu = \mathbf{w}_\nu^T \cdot \lambda_\mu \cdot \mathbf{w}_\mu = \lambda_\mu \cdot \mathbf{w}_\nu^T \cdot \mathbf{w}_\mu \quad \text{(A)} \\ \mathbf{w}_\mu^T \cdot \mathbf{C}\cdot\mathbf{w}_\nu = \mathbf{w}_\mu^T \cdot \lambda_\nu \cdot \mathbf{w}_\nu = \lambda_\nu \cdot \mathbf{w}_\nu^T \cdot \mathbf{w}_\mu \quad \text{(B)}\end{array}$$

$$\mathbf{w}_\nu^T \cdot \mathbf{C}\cdot\mathbf{w}_\mu = \left(\mathbf{w}_\nu^T \cdot \mathbf{C}\cdot\mathbf{w}_\mu\right)^T = \mathbf{w}_\mu^T \cdot \mathbf{C}^T \cdot \mathbf{w}_\nu = \mathbf{w}_\mu^T \cdot \mathbf{C}\cdot\mathbf{w}_\nu \quad (\Gamma)$$

Υπενθυμίζεται ότι $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Από τις σχέσεις (A), (B), (Γ) συνεπάγεται

$$\lambda_\mu \cdot \mathbf{w}_\nu^T \cdot \mathbf{w}_\mu = \lambda_\nu \cdot \mathbf{w}_\nu^T \cdot \mathbf{w}_\mu \Leftrightarrow \left(\lambda_\mu - \lambda_\nu\right)\mathbf{w}_\nu^T \cdot \mathbf{w}_\mu = 0$$

επειδή $\lambda_\mu \neq \lambda_\nu$ έπεται ότι $\mathbf{w}_\nu^T \cdot \mathbf{w}_\mu = \mathbf{w}_\mu^T \cdot \mathbf{w}_\nu = 0$.

Από τα προηγούμενα συνεπάγεται ότι για τον πίνακα $\mathbf{W} = [\mathbf{w}_1, \quad \mathbf{w}_2, \quad \cdots, \quad \mathbf{w}_N]$ ισχύει ότι

$$\mathbf{W}^T \cdot \mathbf{W} = \begin{bmatrix}\mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_N^T\end{bmatrix}[\mathbf{w}_1, \quad \mathbf{w}_2, \quad \cdots, \quad \mathbf{w}_N] = \begin{bmatrix}\mathbf{w}_1^T \cdot \mathbf{w}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{w}_2^T \cdot \mathbf{w}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{w}_N^T \cdot \mathbf{w}_N\end{bmatrix} = \begin{bmatrix}1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 1\end{bmatrix} = \mathbf{I}$$

Άρα αντίστροφος του $\mathbf{W}$ είναι ο $\mathbf{W}^{-1} = \mathbf{W}^T$.

Για τα ιδιοδιανύσματα του $\mathbf{C}$ ισχύει εξ' ορισμού ότι

$$\left.\begin{array}{l}\mathbf{C}\cdot\mathbf{w}_1 = \lambda_1 \cdot \mathbf{w}_1 \\ \mathbf{C}\cdot\mathbf{w}_2 = \lambda_2 \cdot \mathbf{w}_2 \\ \vdots \\ \mathbf{C}\cdot\mathbf{w}_N = \lambda_N \cdot \mathbf{w}_N\end{array}\right\} \Rightarrow \mathbf{C}\cdot[\mathbf{w}_1, \quad \cdots, \quad \mathbf{w}_N] = [\mathbf{w}_1, \quad \cdots, \quad \mathbf{w}_N]\begin{bmatrix}\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N\end{bmatrix} = \mathbf{W}\cdot\mathbf{\Lambda}$$

όπου $\mathbf{\Lambda} = \begin{bmatrix}\lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N\end{bmatrix}$

Αφού $\mathbf{C}\cdot\mathbf{W} = \mathbf{W}\cdot\mathbf{\Lambda} \Rightarrow \mathbf{W}^{-1}\cdot\mathbf{C}\cdot\mathbf{W} = \mathbf{W}^{-1}\cdot\mathbf{W}\cdot\mathbf{\Lambda} = \mathbf{I}\cdot\mathbf{\Lambda} = \mathbf{\Lambda} \Leftrightarrow \mathbf{W}^T\cdot\mathbf{C}\cdot\mathbf{W} = \mathbf{\Lambda}$.