ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Εφαρμογες Της Ψηφιακης Επεξεργασιας Σηματων

Κατηγοριοποιηση & Αναγνωριση Προσωπων

EigenFaces-FisherFaces

Διδάσκων: Αναπλ. Καθηγητής Εμμανουήλ Ζ. Ψαράκης
Επικουρικό έργο: Παναγιώτης Γεωργαντόπουλος

Πάτρα Οκτώβριος 2021

Γραμμική Διαχωριστική Ανάλυση

Η Ευκλείδεια απόσταση είναι η απλούστερη και πιο συχνά χρησιμοποιούμενη μετρική που χρησιμοποιείται για την ταξινόμηση δεδομένων. Η βασική υπόθεση είναι ότι μπορούμε να ορίσουμε ένα αντιπροσωπευτικό μέλος για κάθε κλάση που εμπλέκεται στο πρόβλημα ταξινόμησης. Συνήθως αυτό το αντιπροσωπευτικό μέλος της κλάσης είναι η μέση τιμή. Στη συνέχεια, κάθε δείγμα δοκιμής εκχωρείται στην κλάση του αντιπροσωπευτικού μέλους που είναι πιο κοντά. Ωστόσο, οι κλάσεις μπορεί να έχουν σημαντική επικάλυψη στο χώρο εισόδου καθιστώντας τη μέτρηση της Ευκλείδειας απόστασης αναποτελεσματική. Για την αντιμετώπιση αυτού του προβλήματος, τα δεδομένα μεταφέρονται ή καλύτερα απεικονίζονται σε έναν άλλο χώρο, το χώρο των χαρακτηριστικών στον οποίο οι κλάσεις είναι καλύτερα διαχωρισμένες. Το Σχήμα 1 δείχνει μια τυπική ακολουθία βημάτων που χρησιμοποιούνται στην υλοποίηση ενός ταξινομητή. Οι τεχνικές μοντελοποίησης σήματος που συζητούνται σε αυτό το άρθρο βοηθούν στον καθορισμό του νέου χώρου χαρακτηριστικών.

Η Γραμμική Διαχωριστική Ανάλυση (ΓΔΑ) (Linear Discriminant Analysis (LDA)) [1], όπως και η Ανάλυση Κύριων Συνιστωσων (Principal Components Analysis (PCA)) [2], είναι ένα εργαλείο για ταξινόμηση δεδομένων πολλών ομάδων και μείωση της διαστατικότητας των δεδομένων. Ουσιασικά, η ΓΔΑ μεγιστοποιεί την αναλογία της διακύμανσης μεταξύ των κατηγοριών προς το διακύμανση εντός της κατηγορίας σε οποιοδήποτε συγκεκριμένο σύνολο δεδομένων, εξασφαλίζοντας έτσι τη μέγιστη διαχωριστικότητα. Η κύρια διαφορά μεταξύ ΓΔΑ και ΑΚΣ είναι ότι η ΑΚΣ κάνει ταξινόμηση χαρακτηριστικών ενώ η ΓΔΑ κάνει ταξινόμηση δεδομένων. Η ΑΚΣ αλλάζει τόσο το σχήμα όσο και τη θέση των δεδομένων στον μετασχηματισμένο χώρο της, ενώ η ΓΔΑ δημιουργεί μια συγκεκριμένη περιοχή απόφασης μεταξύ των κλάσεων [3].

Ας υποθέσουμε ότι θέλουμε να διαχωρίσουμε δύο κατηγορίες δεδομένων οι οποίες μπορούν να ταυτοποιηθούν από δύο Γκαουσιανές διανυσματικές τυχαίες μεταβλητές στο χώρο $\mathbb{R}^N$, δηλαδή:

$$\begin{aligned}\mathcal{X}_+ &\sim \mathcal{N}(\mathbf{x}; \mu_+, \Sigma_+)\\ \mathcal{X}_- &\sim \mathcal{N}(\mathbf{x}; \mu_-, \Sigma_-)\end{aligned}$$

όπου $\mu_\pm, \Sigma_\pm)$ δηλώνουν το μέσο όρο και τη συνδιασπορά των τ.μ. που ταυτοποιούν τη θετική και την αρνητική κλάση αντίστοιχα.

Η γραμμική διαχωριστική ανάλυση Fisher (Fisher Linear Discriminant Analysis (FLDA)) είναι μια ευρέως χρησιμοποιούμενη τεχνική για την επίλυση του παραπάνω προβλήματος, και προτάθηκε από τον R. Fisher τη δεκαετία του 1930 [4].

Σύμφωνα με την τεχνική αυτή, θέλουμε να λύσουμε το παραπάνω πρόβλημα επιλέγοντας ένα $\mathbf{w} \in \mathbb{R}^N$, που θα μεγιστοποιεί την παρακάτω πιθανότητα:

$$\max_{\mathbf{w}} \mathbb{P}(\mathbf{w}^t \mathcal{X}_+ > \mathbf{w}^t \mathcal{X}_-) \tag{1}$$

ή ισοδύναμα:

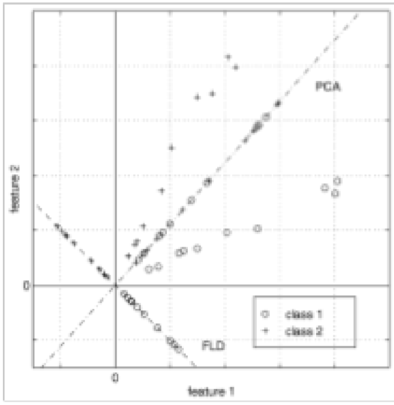
$$\max_{\mathbf{w}} \mathbb{F}_{\mathcal{Z}}(z > 0) \tag{2}$$

όπου $\mathcal{Z} = \mathbf{w}^t(\mathcal{X}_+ - \mathcal{X}_-)$ τ.μ. της οποίας την σ.π.π, θα υπολογίσουμε στην συνέχεια και $\mathbb{F}_{\mathcal{Z}}(z)$ η συνάρτηση κατανομής της.

Για το σκοπό αυτό θα χρειαστούμε τα ακόλουθα:

- Ας υποθέσουμε ότι $\mathbf{w} \in \mathbb{R}^N$ και:

$$\mathcal{X} \sim \mathcal{N}(\mathbf{x}; \mu, \Sigma)$$



Σχήμα 1: LDA vs PCA

μία κανονική διανυσματική τ.μ. με μέση τιμή μ και μητρώο συνδιασπορών Σ. Τότε, για την τ.μ.:

$$\mathcal{Y} =< \mathbf{w},\ \mathcal{X} >= \mathbf{w}^t \mathcal{X}$$

μπορούμε να αποδείξουμε τα ακόλουθα:

$$\mathcal{Y} \sim \mathcal{N}(y; < \mathbf{w},\ \mu >, \mathbf{w}^t \Sigma \mathbf{w})$$

και

$$\mathbb{P}(\{\mathcal{Y}(\theta) > 0\}) = \Phi\Big(\frac{< \mathbf{w},\ \mu >}{(\mathbf{w}^t \Sigma \mathbf{w})^{\frac{1}{2}}}\Big) \tag{3}$$

όπου $\Phi(.)$ η συνάρτηση κατανομής της κανονικής σ.π.π..

- Επιπλέον, αν:

$$\mathcal{X}_i \sim \mathcal{N}(\mathbf{x}_i; \mu_i, \Sigma_i),\ i = 1, 2$$

δύο τ.μ. τότε για την τ.μ.:

$$\mathcal{Y} = \mathcal{X}_1 - \mathcal{X}_2$$

θα έχουμε:

$$\mathcal{Y} \sim \mathcal{N}(\mathbf{x}; \mu_1 - \mu_2, \Sigma_1 + \Sigma_2).$$

Συνδιάζοντας τα παραπάνω και την μονοτονία κάθε σ.π.π., εύκολα βρίσκουμε ότι η λύση του αρχικού προβλήματος βελτιστοποίησης θα είναι η ακόλουθη:

$$\mathbf{w}^\star = (\Sigma_+ - \Sigma_-)^{-1}(\mu_+ - \mu_-). \tag{4}$$

με βέλτιστη τιμή την $(\mu_+ - \mu_-)^t(\Sigma_+ - \Sigma_-)^{-1}(\mu_+ - \mu_-)$.

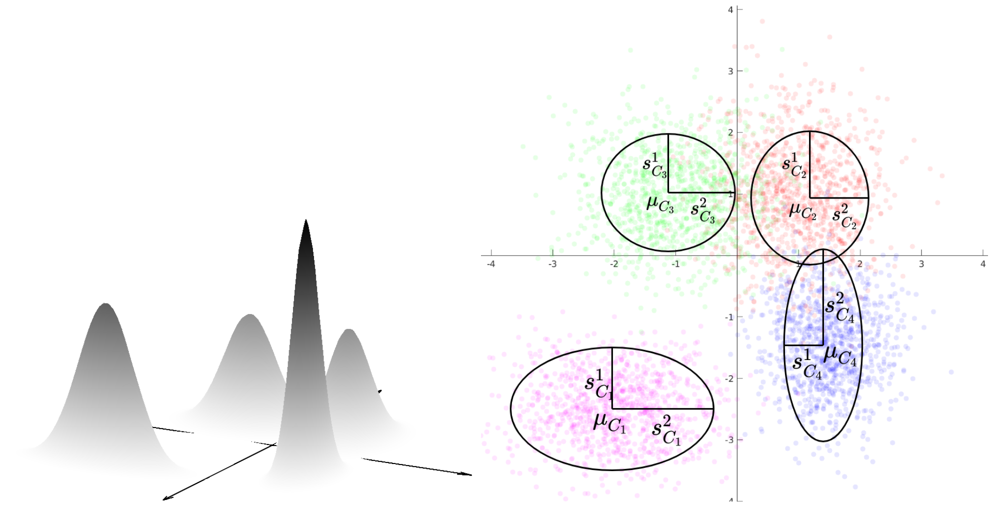
Αν έχουμε βρεί την βέλτιστη τιμή $\mathbf{w}^\star$, μπορούμε να χρησιμοποιήσουμε τον ακόλουθο δυαδικό (binary) κατηγοριοποιητή:

$$\phi(\mathbf{x}) = sign\big(< \mathbf{w}^\star,\ \mathbf{x} > \big) \tag{5}$$

όπου

Ας δούμε τώρα πως μπορούμε να γενικεύσουμε την ιδέα αυτή σε περισσότερες από δύο ομάδες δεδομένων. Για το σκοπό αυτό ας υποθέσουμε ότι έχουμε στη διάθεση μας, σε μορφή ακολουθίας μητρώων, το ακόλουθο σύνολο δεδομένων:

$$X = \{X_j\}_{j=1}^{N_C} \tag{6}$$



Σχήμα 2: Το μοντέλο των συναρτήσεων πυκνότητας πιθανότητας σύμφωνα με το οποίο θεωρούμε ότι κατανέμονται τα δεδομένα μας (αριστερά). Οπτικοποίηση των παραμέτρων που λαμβάνει υπόψιν η LDA στην μοντελοποίηση των δεδομένων (δεξιά).

στο οποίο περιέχονται N_C διαφορετικές ομάδες, ή κατηγορίες, αντικειμένων και X_j ένα $M \times N_j$ μητρώο με το N_j να δηλώνει το πλήθος των δειγμάτων των δεδομένων που ανήκουν στην j-οστή κατηγορία, δηλαδή:

$$X_j = [\mathbf{x}_1^j \; \mathbf{x}_2^j \; \cdots \; \mathbf{x}_{N_j}^j], \; j = 1, 2, \cdots, N_C. \tag{7}$$

Το μητρώο συνδιασπορών S_j της κατηγορίας $\mathcal{C}_j$ ορίζεται ως ακολούθως:

$$S_j = \sum_{i=1}^{|\mathcal{C}_j|} (\mathbf{x}_i^j - \mu_{\mathcal{C}_j})(\mathbf{x}_i^j - \mu_{\mathcal{C}_j})^t \tag{8}$$

όπου $\mu_{\mathcal{C}_j}$ το μέσο διάνυσμα της κατηγορίας $\mathcal{C}_j$, δηλαδή:

$$\mu_{\mathcal{C}_j} = \frac{1}{|\mathcal{C}_j|} \sum_{i=1}^{|\mathcal{C}_j|} \mathbf{x}_i^j, \tag{9}$$

$\mathbf{x}_i^j$ η i-οστη στήλη του μητρώου $\mathcal{X}_j$ που περιέχει τα δείγματα της κατηγορίας C_j καί $|\mathcal{C}_j|$ ο πληθικός αριθμός του συνόλου $\mathcal{C}_j$.

Το μητρώο ενδο-κατηγοριακής διασποράς μίας ομάδας, που συμβολίζουμε με S_{intra}, αποτυπώνει την μέση διασπορά του συνόλου των N_C κατηγοριών και ορίζεται ως ακολούθως:

$$S_{intra} = \frac{1}{N_C} \sum_{j=1}^{N_C} S_j. \tag{10}$$

Ουσιαστικά αυτό το μητρώο (βλ. Σχήμα (1)) αποτυπώνει στο παράδειγμά μας την ακτίνα του μέσου κύκλου.

Το μητρώο δια-κατηγοριακής διασποράς S_{inter} μετρά την διασπορά μεταξύ των κατηγοριών και ορίζεται ως:

$$S_{inter} = \sum_{j=1}^{N_C} |C_j| (\mu_{C_j} - \mu_X)(\mu_{C_j} - \mu_X)^t \tag{11}$$

όπου μ_X το μέσο διάνυσμα του μητρώου δεδομένων X, δηλαδή:

$$\mu_X = \frac{1}{N_C}\sum_{j=1}^{N_C}\mu_{C_j}. \tag{12}$$

Το μητρώο:

$$S = S_{intra}^{-1}S_{inter} \tag{13}$$

μπορεί να χρησιμοποιηθεί για τη μέτρηση της διαχωρισμού των ομάδων, εφαρμόζοντας Ανάλυση Κυρίων Συνιστωσών (PCA).

Ο Αλγόριθμος K-Means

Δίνοντας έμφαση στην ενδο-κατηγοριακή διασπορά της Ενότητας 2, θα παρουσιάσουμε τον αλγόριθμο K-means [5]. Αρχικά, ωστόσο, ας ορίσουμε κάποια βασικά μεγέθη.

Για το σκοπό αυτό, ας θεωρήσουμε δύο διανύσματα του χώρου R^N, δηλαδή:

$$\begin{aligned}\mathbf{x} &= \begin{bmatrix}x_1 & x_2 & \dots & x_N\end{bmatrix}^t \\ \mathbf{y} &= \begin{bmatrix}y_1 & y_2 & \dots & y_N\end{bmatrix}^t.\end{aligned} \tag{14}$$

Τότε:

- η Ευκλείδια, ή l_2 στάθμη, μεταξύ των παραπάνω διανυσμάτων ορίζεται ως:

$$\|y-x\|_2 = \sqrt{\sum_{n=1}^{N}(y_n - x_n)^2}. \tag{15}$$

- η l_1 νόρμα μεταξύ των ορίζεται ως:

$$\|y-x\|_1 = \sum_{n=1}^{N}|y_n - x_n|, \tag{16}$$

- ενώ η στάθμη l_∞ ορίζεται ως:

$$\|y-x\|_\infty = \max_{1\le n\le N}\{|y_n - x_n|\}. \tag{17}$$

Ας υποθέσουμε τώρα ότι γνωρίζουμε τα ακόλουθα σύνολα δεικτών:

$$\mathcal{E}_k = \{i_1^k, i_2^k, \cdots i_{n_k}^k\},\ k = 1,2,\dots,K \text{ με } i_j^k \in \mathbb{Z}. \tag{18}$$

Προφανώς, ισχύει $|\mathcal{E}_k| = n_k$ με $\sum\limits_k n_k = N$.

Ας ορίσουμε τώρα την ακόλουθη συνάρτηση κόστους:

$$\mathcal{G}_2(\mathbf{z}_k) = \sum_{j=1}^{n_k}\left\|\mathbf{x}_{i_j^k} - \mathbf{z}_k\right\|_2^2,\ k = 1,2,\dots,K. \tag{19}$$

Εάν γνωρίζουμε τα $\mathbf{x}_{i_j^k}$, $j = 1, 2, \cdots, n_k$, $k = 1, 2, \ldots, K$, τότε είναι εύκολο να υπολογίσουμε τα βέλτιστα $\mathbf{z}_k$, $k = 1, 2, \cdots, K$:

$$\mathbf{z}_k^* = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{x}_{i_j^k} \tag{20}$$

όπως εύκολα προκύπτει από την ελαχιστοποίηση των παραπάνω συνάρτησεων κόστους, ήτοι:

$$\mathbf{z}_k^* = \arg\min_{\mathbf{z}_k} \{\mathcal{G}_2(\mathbf{z}_k)\}, \; k = 1, 2, \ldots, K. \tag{21}$$

Θα πρέπει στο σημείο αυτό να τονίσουμε το γεγονός ότι η υπόθεση που κάναμε σχετικά με την γνώση των συνόλων δεικτών της Σχέσης (18) δεν ισχύει. Ως εκ τούτου στην πραγματικότητα θα ορίσουμε την παρακάτω συνολική συνάρτηση κόστους:

$$\mathcal{C}(\{\mathcal{E}_k, \; \mathbf{z}_k\}_{k=1}^K) = \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^{n_k} \left\| \mathbf{x}_{i_j^k} - \mathbf{z}_k \right\|_2^2 \tag{22}$$

την οποία θέλουμε να ελαχιστοποιήσουμε, δηλαδή:

$$\min_{\{\mathcal{E}_k, \; \mathbf{z}_k\}_{k=1}^K} \mathcal{C}(\{\mathcal{E}_k, \; \mathbf{z}_k\}_{k=1}^K). \tag{23}$$

Αν και η ελαχιστοποίηση ως προς $\mathbf{z}_k$, όπως είδαμε παραπάνω, είναι τετριμμένη όταν είναι γνωστά τα σύνολα $\mathcal{E}_k$, $k = 1, 2, \cdots, K$, αυτά δεν είναι γνωστά και ο υπολογισμός των προκύπτει επαναληπτικά. Δηλαδή, σχηματίζουμε δύο προβλήματα που μπορούν να επιλυθούν ακολουθιακά, καθώς το ένα θεωρεί την λύση του άλλου γνωστή. Για το σκοπό αυτό στην $l = 1$ επαναληψη του αλγορίθμου $k - means$, επιλέγουμε K διανύσματα από τα δεδομένα με τυχαίο τρόπο, ήτοι:

$$\mathcal{P}_1 : \text{Επίλεξε τυχαία } \mathbf{z}_k^l, \quad k = 1, \ldots, K. \tag{24}$$

Έχοντας τα κέντρα των κατηγοριών, μπορούμε να ορίσουμε τα σύνολα $\mathcal{E}_k, \quad k = 1, \ldots, K$, από την λύση του ακόλουθου προβλήματος για κάθε δείγμα $\mathbf{x}_j$ του συνόλου των δεδομένων:

$$\mathcal{P}_2 : \mathbf{x}_j \in \mathcal{E}_{k^\star} \text{με } k^\star = \arg\min_k ||\mathbf{z}_k - \mathbf{x}_j||_2^2. \tag{25}$$

Έχοντας δημιουργήσει τα παραπάνω σύνολα, μπορούμε να χρησιμοποιήσουμε τη Σχέση (20) και να υπολογίσουμε εκ νέου τα κέντρα $\mathbf{z}_k^l$, $\;k = 1, \ldots, K$ και η διαδικασία να επαναλαμβάνεται μέχρι να τηρηθεί η συνθήκη:

$$\sum_{k=1}^K \left\| \mathbf{z}_k^l - \mathbf{z}_k^{l-1} \right\|_2^2 \leq \epsilon, \tag{26}$$

όπου ϵ προεπιλεγμένη σταθερά.

Η μόνη ποσότητα την οποία συνήθως δεν γνωρίζουμε εκ των προτέρων, είναι το πλήθος των κλάσεων των αντικειμένων που υπάρχουν στα δεδομένα. Είναι προφανές ότι μπορούμε να εκχωρίσουμε κάθε δείγμα σε μία κατηγορία. Ωστόσο, αυτό δεν μας δινει την συνολική εικόνα για την διασπορά των δειγμάτων και κατά πόσο αυτή οφείλεται σε μία ή περισσότερες κλάσεις. Οπότε, πρέπει να επιλέξουμε ένα πλήθος κατηγοριών K έτσι ώστε να καταλήξουμε σε ένα μοντέλο που να μην είναι, όπως καλούμε, υπερπροσαρμοσμένο στα δεδομένα μας.

Σε αυτή την εξισορρόπηση μεταξύ υπερπαραμέτρων ενός συστήματος (στην περίπτωσή μας το K), μπορούμε να χρησιμοποιήσουμε το Κριτήριο Πληροφορίας του Akaike (AIC) ή το Μπεϋζιανό Κριτήριο Πληροφορίας (BIC).

Ο Αλγόριθμος K-means για την στάθμη l_1

Είδαμε πως ο K-means λύνει το πρόβλημα χρησιμοποιώντας την στάθμη l_2. Ας εξετάσουμε τώρα πως επιδρά στη λύση του αλγορίθμου η χρήση της στάθμης l_1. Συγκεκριμένα, θα πρέπει να ορίσουμε το αντίστοιχο της συνάρτησης κόστους της Σχέσης (19) για την στάθμη l_1, δηλαδή:

$$\mathcal{G}_1(\mathbf{z}_k) = \sum_{j=1}^{n_k} \left\| \mathbf{x}_{i_j^k} - \mathbf{z}_k \right\|_1, \ k = 1, 2, \dots, K \tag{27}$$

και για κάθε σύνολο $\mathcal{E}_k, \ k = 1, 2, \cdots, K$, να λύσουμε το αντίστοιχο πρόβλημα ελαχιστοποίησης, ήτοι:

$$\mathbf{z}_k^\star = \arg\min_{\mathbf{z}_k} \mathcal{G}_1(\mathbf{z}_k) \tag{28}$$

του οποίου η βέλτιστη λύση δίνεται από το μεσαίο (median) διάνυσμα της κλάσης, ήτοι:

$$\mathbf{z}_k^\star = median\{\mathbf{x}_{1k}, \mathbf{x}_{2k}, \dots, \mathbf{x}_{n_k k}\}, \quad k = 1, 2, \dots, K. \tag{29}$$

Canonical Correlation Analysis

Θα παρουσιάσουμε τώρα το πρόβλημα του Canonical Correlation Analysis [6]. Ας θεωρήσουμε δύο διανυσματικές τυχαίες μεταβλητές[1] $\mathcal{X}$ και $\mathcal{Y}$. Η συνδιασπορά τους ορίζεται ως:

$$\Sigma_{\mathcal{X}\mathcal{Y}} = \mathbb{E}\{\mathcal{X}\mathcal{Y}^t\}. \tag{30}$$

Ας υποθέσουμε τώρα ότι ορίζουμε με την βοήθεια των μητρώωνν $W_i, \ i = 1, 2$ τις ακόλουθες τ. μ.:

$$\begin{aligned} \mathcal{X}^{'} &= W_1\mathcal{X} \quad \text{και} \\ \mathcal{Y}^{'} &= W_2\mathcal{Y}. \end{aligned} \tag{31}$$

Το πρόβλημα που θέλουμε να λύσουμε μπορεί να διατυπωθεί ως ακολούθως.

Αναζητούμε εκείνα τα μητρώα $W_i, \ i = 1, 2$ τα οποία μεγιστοποιούν το μητρώο ετεροσυσχέτισης $\Sigma_{\mathcal{X}'\mathcal{Y}'}$ των παραπάνω τ.μ., το οποίο ορίζεται ως:

$$\Sigma_{\mathcal{X}'\mathcal{Y}'} = W_1 \Sigma_{\mathcal{X}\mathcal{Y}} W_2^T. \tag{32}$$

Σε αυτό το σημείο είναι χρήσιμη η ακόλουθη ανισότητα η οποία είναι γνωστή ως ανισότητα των Cauchy-Schwartz:

- Περίπτωση 1: Τα $\mathbf{x}, \ \mathbf{y}$ είναι διανύσματα του χώρου $\mathbb{R}^d, \ d \in \ \mathbb{Z}$. Σε αυτή την περίπτωση η ανισότητα εκφράζεται από το μέτρο του ακόλουθου εσωτερικού γινομένου:

$$|<\mathbf{x}, \ \mathbf{y}>| = |\mathbf{x}^t\mathbf{y}| \leq ||\mathbf{x}||_2||\mathbf{y}||_2, \tag{33}$$

  με ικανή και αναγκαία συνθήκη για να ισχύει η ισότητα $\mathbf{x} = \alpha\mathbf{y}$ (αποδείξτε το).

[1] Όπου θα αναφερόμαστε σε τ.μ. θα θεωρούμε, αν δεν αναφέρεται κάτι άλλο, ότι αυτές είναι μηδενικής μέσης τιμής.

- Περίπτωση 2: Τα $\mathcal{X}, \mathcal{Y}$ είναι διανυσματικές τ.μ. της ίδιας κατανομής. Σε αυτή την περίπτωση η ανισότητα εκφράζεται από το μέτρο του ακόλουθου εσωτερικού γινομένου:

$$\left|\mathbb{E}[< \mathcal{X},\ \mathcal{Y} >]\right| \leq \sqrt{\mathbb{E}[||\mathcal{X}||_2^2]}\sqrt{\mathbb{E}[||\mathcal{Y}||_2^2]}, \tag{34}$$

με ικανή και αναγκαία συνθήκη για να ισχύει η ισότητα $\mathcal{X} = \alpha\mathcal{Y}$ (αποδείξτε το).

Μια άλλη χρήσιμη ιδιότητα που θα μας χρειαστεί στην λύση του προβλήματος που διατυπώσαμε παραπάνω, είναι η ακόλουθη:

Εάν ένα μητρώο W είναι συμμετρικό και θετικά λορισμένο, τότε αυτό μπορεί να διαγωνοποιηθεί ως εξής:

$$W = V\Lambda V^T, \tag{35}$$

όπου V ένα ορθοκανονικό μητρώο:

$$V^T V = VV^T = I$$

του οποίου η j-οστή στήλη $\mathbf{v}_j$ είναι το ιδιοδιάνυσμα του μητρώου που αντιστοιχεί στην j-οστή ιδιοτιμή του λ_j και L ένα διαγώνιο μητρώο που περιέχει τις ιδιοτιμές του μητρώου W, δηλαδή:

$$L = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix}, \quad \lambda_i \in \mathbb{R} \quad \text{και} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N. \tag{36}$$

Θα πρέπει να τονίσουμε στο σημείο αυτό, ότι η παραπάνω ιδιότητα έχει ήδη χρησιμοποιηθεί σε προηγούμενη άσκηση του μαθήματος για τη λύση του παρακάτω, υπό συνθήκη προβλήματος βελτιστοποίησης:

$$\max_{\mathbf{x} \in\ \mathbb{R}^d} \mathbf{x}^t W \mathbf{x}, \quad \text{με συνθήκη } \|\mathbf{x}\|_2 = 1. \tag{37}$$

Η λύση του προβλήματος (37) είναι η:

$$\mathbf{x}^* = \mathbf{v}_1 \quad \text{με} \quad \mathbf{x}^{*t} W \mathbf{x}^* = \lambda_1, \tag{38}$$

όπου $\mathbf{v}_1$ το ιδιοδιάνυσμα του μητρώου W που αντιστοιχεί στην μέγιστη ιδιοτιμή του λ_1.

Απόδειξη: Το πρόβλημα (37) είναι ισοδύναμο με το ακόλουθο:

$$\max_{\mathbf{x} \in\ \mathbb{R}^d} \frac{\mathbf{x}^t W \mathbf{x}}{\|\mathbf{x}\|_2}. \tag{39}$$

Χρησιμοποιώντας την Σχέση (35), η παραπάνω σχέση γράφεται ισοδύναμα ως εξής:

$$\frac{\mathbf{x}^t W \mathbf{x}}{\|\mathbf{x}\|_2} = \frac{\mathbf{x}^t V \Lambda V^T \mathbf{x}}{\|\mathbf{x}\|_2}. \tag{40}$$

Παίρνοντας υπόψην μας τώρα την ορθοκανονικότητα του μητρώου V τη μορφή και το περιεχόμενο του μητρώου Λ, η λύση του προβλήματος όπως αυτή δίνεται στη Σχέση (38) προκύπτει εύκολα. *QED.*

Ας επιστρέψουμε τώρα στο αρχικό πρόβλημα και ας εξετάσουμε αρχικά την απλή περίπτωση που τα $\mathcal{X}'$ και $\mathcal{Y}'$ είναι βαθμωτές τ.μ.. Σε αυτή την περίπτωση η Σχέση (31) γράφεται ως ακολούθως:

$$\mathcal{X}' = \mathbf{w}_1^t \mathcal{X} \quad \text{και} \tag{41}$$

$$\mathcal{Y}' = \mathbf{w}_2^t \mathcal{Y} \tag{42}$$

με $\mathbf{w}_i, \ i = 1, 2$ διανύσματα.

Ο συντελεστής συσχέτισης των παραπάνω τ.μ. ορίζεται από την ακόλουθη σχέση:

$$\rho_{\mathcal{X}'\mathcal{Y}'} = \frac{\mathbb{E}[\mathcal{X}'\mathcal{Y}']}{\sqrt{\mathbb{E}[\mathcal{X}'^2]}\sqrt{\mathbb{E}[\mathcal{Y}'^2]}}, \tag{43}$$

με $|\rho_{\mathcal{X}'\mathcal{Y}'}| \leq 1$.

Χρησιμοποιώντας τις Σχέσεις (30), (42), η Σχέση (43) ισοδύναμα μπορεί να εκφραστεί ως ακολούθως:

$$\begin{aligned} \rho_{\mathcal{X}'\mathcal{Y}'} &= \frac{\mathbb{E}[\mathbf{w_1}^t \mathcal{X}\mathcal{Y}^t \mathbf{w_2}]}{\sqrt{\mathbb{E}[\mathbf{w_1}^t \mathcal{X}\mathcal{X}^t \mathbf{w_1}]}\sqrt{\mathbb{E}[\mathbf{w_2}^t \mathcal{Y}\mathcal{Y}^t \mathbf{w_2}]}} \\ &= \frac{\mathbf{w_1}^t \mathbb{E}[\mathcal{X}\mathcal{Y}^t]\mathbf{w_2}}{\sqrt{\mathbf{w_1}^t \mathbb{E}[\mathcal{X}\mathcal{X}^t]\mathbf{w_1}}\sqrt{\mathbf{w_2}^t \mathbb{E}[\mathcal{Y}\mathcal{Y}^t]\mathbf{w_2}}} \\ &= \frac{\mathbf{w_1}^t \Sigma_{\mathcal{X}\mathcal{Y}} \mathbf{w_2}}{\sqrt{\mathbf{w_1}^t \Sigma_{\mathcal{X}\mathcal{X}} \mathbf{w_1}}\sqrt{\mathbf{w_2}^t \Sigma_{\mathcal{Y}\mathcal{Y}} \mathbf{w_2}}}. \end{aligned} \tag{44}$$

Όπως είπαμε θέλουμε να μεγιστοποιήσουμε την παραπάνω ποσότητα. Για το σκοπό αυτό, ας ορίσουμε τις ακόλουθες ντετερμινιστικές ποσότητες:

$$\begin{aligned} \mathbf{x}_{\mathcal{X}} &= \Sigma_{\mathcal{X}\mathcal{X}}^{\frac{1}{2}} \mathbf{w_1} \\ \mathbf{y}_{\mathcal{Y}} &= \Sigma_{\mathcal{Y}\mathcal{Y}}^{\frac{1}{2}} \mathbf{w_2}, \end{aligned} \tag{45}$$

όπου $A^{\frac{1}{2}}$ η τετραγωνική ρίζα του θετικά ορισμένου μητρώου A.

Αντικαθιστώντας τις ποσότητες της Σχέσης (45) στη Σχέση (44) έχουμε:

$$\rho_{\mathcal{X}'\mathcal{Y}'} = \frac{\mathbf{x}_{\mathcal{X}}^t \Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} \mathbf{y}_{\mathcal{Y}}}{||\mathbf{x}_{\mathcal{X}}||_2 ||\mathbf{y}_{\mathcal{Y}}||_2}. \tag{46}$$

Χρησιμοποιώντας τώρα την ανισότητα των Cauchy-Schwartz, από την παραπάνω σχέση παίρνουμε:

$$\begin{aligned} \rho_{\mathcal{X}'\mathcal{Y}'} &\leq \frac{||\mathbf{x}_{\mathcal{X}}||_2 ||\Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} \mathbf{y}_{\mathcal{Y}}||_2}{||\mathbf{x}_{\mathcal{X}}||_2 ||\mathbf{y}_{\mathcal{Y}}||_2} \\ &= \frac{||\Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} \mathbf{y}_{\mathcal{Y}}||_2}{||\mathbf{y}_{\mathcal{Y}}||_2} \\ &= \frac{||\Sigma \mathbf{y}_{\mathcal{Y}}||_2}{||\mathbf{y}_{\mathcal{Y}}||_2}, \end{aligned} \tag{47}$$

όπου:

$$\Sigma = \Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} \tag{48}$$

και με την ισότητα να ισχύει αν και μόνο αν:

$$\mathbf{x}_{\mathcal{X}} = \alpha \Sigma_{\mathcal{X}\mathcal{X}}^{\frac{-1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{T\frac{-1}{2}} \mathbf{y}_{\mathcal{Y}}. \tag{49}$$

Όπως ήδη είπαμε το μητρώο Σ είναι θετικά ορισμένο και επομένως η συνάρτηση ομοιότητας που ορίσαμε παραπάνω, είναι ισοδύναμη με την ακόλουθη:

$$\frac{||\Sigma \mathbf{y}_{\mathcal{Y}}||_2}{||\mathbf{y}_{\mathcal{Y}}||_2} \sim \frac{\mathbf{y}_{\mathcal{Y}}^t \Sigma^T \Sigma \mathbf{y}_{\mathcal{Y}}}{||\mathbf{y}_{\mathcal{Y}}||_2^2}, \tag{50}$$

της οποίας η μεγιστοποίηση, χρησιμοποιώντας το πρόβλημα βελτιστοποίησης (37), είναι τετριμένη.

Αν συμβολίσουμε με $\mathbf{y}_{\mathcal{Y}}^\star$ την βέλτιστη λύση του παραπάνω προβλήματος, τότε τα βέλτιστα ντετερμινιστικά διανύσματα $\mathbf{w}_i,\ i = 1, 2$ ορίζονται από τις ακόλ,ουθες σχέσεις:

$$\begin{aligned} \mathbf{w}_1 &= \alpha \Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} \mathbf{y}_{\mathcal{Y}}^\star & (51) \\ \mathbf{w}_2 &= \Sigma_{\mathcal{Y}\mathcal{Y}}^{\frac{-1}{2}} \mathbf{y}_{\mathcal{Y}}^\star & (52) \end{aligned}$$

και αυτό ολοκληρώνει τη λύση του προβλήματος.

Στη γενική περίπτωση όπου τα $\mathcal{X}'$, $\mathcal{Y}'$ είναι διανυσματικές τ.μ., όπως αυτά ορίσθηκαν στη Σχέση (31), η βέλτιστη λύση θα δίνεται από τις ακόλουθες σχέσεις:

$$\begin{aligned} W_1 &= \Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} V & (53) \\ W_2 &= \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}} U & (54) \end{aligned}$$

όπου τα μητρώα V, U προκύπτουν από την SVD ανάλυση του:

$$\Sigma_{\mathcal{X}\mathcal{X}}^{-\frac{1}{2}} \Sigma_{\mathcal{X}\mathcal{Y}} \Sigma_{\mathcal{Y}\mathcal{Y}}^{-\frac{1}{2}T} = V \Lambda U^T. \tag{55}$$

Παρατηρείστε ότι όλοι οι απαραίτητοι περιορισμοί τηρούνται. Δηλαδή:

$$\begin{aligned} W_1^T \Sigma_{\mathcal{X}\mathcal{X}} W_1 &= I \\ W_2^T \Sigma_{\mathcal{Y}\mathcal{Y}} W_2 &= I, \text{και} \\ \Sigma_{\mathcal{X}'\mathcal{Y}'} &= \Lambda. \end{aligned}$$

Διαδικασία

1. Σας δίνεται ένα σύνολο εικόνων (βλ. faces_dataset.zip) που απεικονίζουν πρόσωπα.
    - Καλείστε να κατηγοριοποιήσετε τα δεδομένα αυτά σε κλάσεις της επιλογής σας (π.χ. πρόσωπα που φορούν γυαλιά, πρόσωπα με μακρυά μαλλιά, πρόσωπα γυναικεία/ανδρικά κλπ.). Η επιλογή θα είναι δική σας σύμφωνα με παρατηρήσεις που εσείς θα κάνετε πάνω στο σύνολο δεδομένων.
    - Έχοντας επιλέξει τις κλάσεις που επιθυμείτε διαχωρήστε το σύνολο των εικόνων βάσει αυτών βάζοντας τες σε ξεχωριστούς φακέλους, είτε κρατώντας μια λίστα με τα ονόματά τους και το όνομα της κλάσης τους, και δημιουργείστε μια δενδρική δομή ώστε να διευκολυνθείτε στην συνέχεια με τον τρόπο που θα τις χρησιμοποιήσετε παρακάτω.

- Δεδομένου του παραπάνω διαχωρισμού χωρίστε τις εικόνες κάθε κλάσης σε δεδομένα εκπαίδευσης και δεδομένα αξιολόγησης. Τα πρώτα θα τα χρησιμοποιήσετε για να εκπαιδεύσετε το σύστημά σας και τα δεύτερα για να μετρήσετε την επίδοσή κάθε τεχνικής σε δεδομένα που ανήκουν μεν στην ίδια κλάση, αλλά τα οποία δεν έχουν ξαναδεί.
- Για κάθε μία από τις τεχνικές που ανατύχθηκαν παραπάνω, χρησιμοποιείστε τα δεδομένα εκπαίδευσης και εκπαιδεύστε έναν κατηγοριοποιητή. Αναλυτικά, δεδομένων των εικόνων εκπαίδευσης, θα υπολογίσετε τις απαραίτητες ποσότητες που να μετασχηματίζουν κατάλληλα τα δεδομένα σας, όπως αυτό περιγράφεται στην αντίστοιχη θεωρία για την επιτυχή επίλυση του προβλήματος.
- Σχολιάστε την απόδοση κάθε τεχνικής και συγκρίνετε την με τις υπόλοιπες. Για το σκοπό αυτό, χρησιμοποιείστε τον μετασχηματισμό που υπολογίσατερ για κάθε τεχνική από τα δεδομένα εκπαίδευσης και προβάλετε τα δεδομένα αξιολόγησης για να δείτε εάν οι προκύπτουσες προβολές είναι κοντά σε κάποια κλάση από τις υπόλοιπες, διαμορφώνωντας έτσι έναν κανόνα απόφασης. Παρουσιάστε το μητρώο σύγχησης (confusion matrix) για τον συγκεκριμένο κατηγοριοποιητή και σχολιάστε το.

Βιβλιογραφία

[1] Zizhu Fan, Yong Xu, and David Zhang. Local linear discriminant analysis framework using sample neighbors. *IEEE Transactions on Neural Networks*, 22(7):1119–1132, 2011.

[2] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[3] Aleix M Martinez and Avinash C Kak. Pca versus lda. *IEEE transactions on pattern analysis and machine intelligence*, 23(2):228–233, 2001.

[4] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[5] Anil Chaturvedi, Paul E Green, and J Douglas Caroll. K-modes clustering. *Journal of classification*, 18(1):35–55, 2001.

[6] David Weenink. *Canonical correlation analysis*, volume 25. 2003.