Ελλάτωση των διαστάσεων

- Principal Component Analysis (PCA)
- Fisher Linear Discriminant Analysis (LDA)

Οφέλη από την ελλάτωση των διαστάσεων

- Δημιουργία νέων χαρακτηριστικών (feature generation) σε συνέχεια της αρχικής διαδικασίας εξαγωγής και επιλογής χαρακτηριστικών
- Αποφυγή της υπερ-προσαρμογής του μοντέλου μάθησης όταν δεν είναι εφικτό να αυξήσουμε το σύνολο των δεδομένων
- Οπτικοποίηση των δεδομένων (π.χ. μείωση σε 2 ή 3 διαστάσεις)
- Συμπίεση των δεδομένων
- Μείωση πλεονασμού των δεδομένων
- Μείωση θορύβου που μολύνει τα δεδομένα
- Αύξηση της ταχύτητας των υπολογισμών (π.χ. αύξηση της ταχύτητας εκπαίδευσης και αύξηση της ταχύτητας λήψης απόφασης)

Principal Component Analysis (PCA)

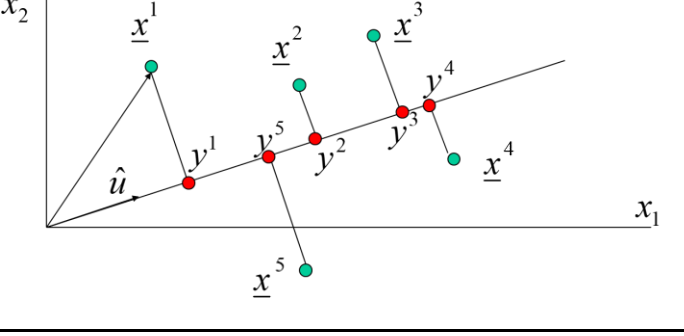
- Η ανάλυση κύριων συνιστωσών (PCA) είναι επίσης γνωστή και ως ο μετασχηματισμός Karhunen-Loeve (KLT).
- Οι συνιστώσες του νέου διανύσματος είναι προβολές του αρχικού στους άξονες που καθορίζονται από τα μοναδιαία διανύσματα $\mathbf{u}_i$.
- ΣΤΟΧΟΙ που πρέπει να ικανοποιηθούν ταυτόχρονα:
- ✓ Οι νέοι άξονες πρέπει να επιλεγούν έτσι ώστε να μεγιστοποιηθεί η διασπορά των χαρακτηριστικών
- ✓ Τα χαρακτηριστικά πρέπει να είναι ανά δύο ασυσχέτιστα (uncorrelated), δηλαδή, αν $\mathrm{E}\{y_i\} = 0 \quad \forall i$, τότε:

$$\mathrm{E}\{y_i y_j\} = 0 \quad \forall i \neq j$$

Principal Component Analysis (PCA)

Η προβολή όλων των διανυσμάτων του εκπαιδευτικού συνόλου σε έναν άξονα που καθορίζεται από το μοναδιαίο διάνυσμα $\mathbf{u}$ γίνεται υπολογίζοντας τα εσωτερικά γινόμενα:

$$y = \mathbf{u}^T \mathbf{x} = \mathbf{x}^T \mathbf{u}$$



Principal Component Analysis (PCA)

- Έστω Σ ο πίνακας συνδιακύμανσης (covariance matrix) που υπολογίζεται από το εκπαιδευτικό σύνολο:

$$\Sigma = E\{(x-\mu)(x-\mu)^T\} = E\{xx^T\} - E\{x\}E\{x^T\}$$

- Μπορεί εύκολα να δειχθεί ότι η προβολή του Σ σε οποιοδήποτε άξονα **u**, δηλαδή η τετραγωνική μορφή $\mathbf{u}^T\Sigma\mathbf{u}$, ισούται με την διασπορά της προβολής (y) του διανύσματος **x** στον ίδιο άξονα:

$$\begin{aligned}\sigma_y^2 &= E\{(y-E\{y\})^2\} = E\{y^2\} - E^2\{y\} = \\ &= E\{(\mathbf{u}^T\mathbf{x})(\mathbf{x}^T\mathbf{u})\} - E\{\mathbf{u}^T\mathbf{x}\}\,E\{\mathbf{x}^T\mathbf{u}\} = \\ &= E\{\mathbf{u}^T(\mathbf{x}\mathbf{x}^T)\mathbf{u}\} - \mathbf{u}^T E\{\mathbf{x}\}E\{\mathbf{x}^T\}\mathbf{u} = \\ &= \mathbf{u}^T E\{\mathbf{x}\mathbf{x}^T\}\mathbf{u} - \mathbf{u}^T E\{\mathbf{x}\}E\{\mathbf{x}^T\}\mathbf{u} = \\ &= \mathbf{u}^T(E\{\mathbf{x}\mathbf{x}^T\} - E\{\mathbf{x}\}E\{\mathbf{x}^T\})\mathbf{u} = \mathbf{u}^T\Sigma\mathbf{u}\end{aligned}$$

Principal Component Analysis (PCA)

1ος Στόχος: Οι άξονες $\mathbf{u}$ ($\mathbf{u}^T\mathbf{u} = 1$) πρέπει να επιλεγούν έτσι ώστε να μεγιστοποιηθεί η διασπορά των χαρακτηριστικών

Σύμφωνα με τα προηγούμενα, θέλουμε να λύσουμε το παρακάτω πρόβλημα βελτιστοποίησης υπό περιορισμούς:

- Μεγιστοποίηση της ποσότητας $\mathbf{u}^T\Sigma\mathbf{u}$ ως προς τους άξονες $\mathbf{u}$ έτσι ώστε $\mathbf{u}^T\mathbf{u} = 1$.
- Δημιουργία της Lagrangian:

$$L = \mathbf{u}^T\Sigma\mathbf{u} - \lambda(\mathbf{u}^T\mathbf{u} - 1) \quad \text{όπου } \lambda \geq 0.$$

- Η λύση μηδενίζει την παράγωγο (κλίση) της L ως προς $\mathbf{u}$:

$$\mathbf{d}L/\mathbf{du} = \nabla_{\mathbf{u}} L = 2\Sigma\mathbf{u} - 2\lambda\mathbf{u} = 0 \implies \Sigma\mathbf{u} = \lambda\mathbf{u}$$

Principal Component Analysis (PCA)

- Συνεπώς, το $\mathbf{u}$ είναι ένα ιδιοδιάνυσμα του πίνακα συνδιακύμανσης.
- Επιπλέον, έχουμε ότι:

$$\Sigma\mathbf{u} = \lambda\mathbf{u} \;\;\Rightarrow\;\; \mathbf{u}^{\mathrm{T}}\Sigma\mathbf{u} = \lambda\mathbf{u}^{\mathrm{T}}\mathbf{u} \;\;\Rightarrow\;\; \mathbf{u}^{\mathrm{T}}\Sigma\mathbf{u} = \lambda$$

  δηλαδή, η διασπορά σ_y^2 της προβολής θα είναι ίση με την ιδιοτιμή που αντιστοιχεί στο ιδιοδιάνυσμα $\mathbf{u}$.

  Επειδή δε, ο πίνακας Σ είναι συμμετρικός και πραγματικός, θα έχει *d* ιδιοδιανύσματα ανά δύο ορθοκανονικά:

$$\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j = \delta_{ij} = \begin{cases} 1 & \text{αν } i = j \\ 0 & \text{αν } i \neq j \end{cases}$$

Principal Component Analysis (PCA)

- Η μέθοδος PCA στοχεύει στην επιλογή των *m* από τα *d* ιδιοδιανύσματα που αντιστοιχούν στις *m* μεγαλύτερες ιδιοτιμές.
- Έστω $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ τα ιδιοδιανύσματα του Σ που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές.
- Η προβολή των *d*-διάστατων διανυσμάτων **x** του εκπαιδευτικού συνόλου στην νέα βάση των *m* ιδιοδιανυσμάτων $\mathbf{u}_i$ (i = 1,...,*m*) τα μετασχηματίζει στα *m*-διάστατα διανύσματα **y** μέσω του παρακάτω γραμμικού μετασχηματισμού:

$$\mathbf{y} = \mathrm{A}^{\mathrm{T}}\mathbf{x}$$

  όπου ο πίνακας A (*d*x*m*) είναι: $\mathrm{A} = [\mathbf{u}_1 | \mathbf{u}_2 | \ldots | \mathbf{u}_m]$.
- Συνεπώς, τα διανύσματα **y** είναι οι προβολές των **x** στον υποχώρο που καλύπτεται από τα *m* κύρια ιδιοδιανύσματα.

Principal Component Analysis (PCA)

2ος Στόχος: Τα χαρακτηριστικά πρέπει να είναι ανά δύο ασυσχέτιστα.

Υποθέτουμε ότι τα αρχικά χαρακτηριστικά έχουν μηδενική μέση τιμή (αυτό είναι εύκολο να επιτευχθεί μετατοπίζοντας κατάλληλα όλο το εκπαιδευτικό σύνολο). Συνεπώς,

$$\Sigma = E\{xx^T\} - E\{x\}E\{x^T\} = E\{xx^T\}$$

και

$$E\{y_i y_j\} = E\{u_i^T xx^T u_j\} = u_i^T E\{xx^T\} u_j = u_i^T \Sigma u_j$$

Όμως, καθώς u_i είναι ένα ιδιοδιάνυσμα του Σ και επειδή τα ιδιοδιανύσματα είναι ορθοκανονικά, θα έχουμε:

$$E\{y_i y_j\} = u_i^T \Sigma u_j = u_i^T \lambda_j u_j = \lambda_j u_i^T u_j = \lambda_j \delta_{ij} = \begin{cases} \lambda_i & \text{αν } i = j \\ 0 & \text{αν } i \neq j \end{cases}$$

Συνεπώς, τα χαρακτηριστικά είναι ασυσχέτιστα.

Principal Component Analysis (PCA)

- Πάλι, βρίσκουμε ότι:

$$E\{y_i^2\} = \lambda_i$$

- Συνεπώς, η διασπορά σε κάθε κύρια κατεύθυνση ισούται με την ιδιοτιμή που αντιστοιχεί στην κατεύθυνση αυτή.

Principal Component Analysis (PCA)

- Τώρα, ας ορίσουμε το

$$\underline{\hat{x}} = \sum_{i=1}^{m} y_i \hat{u}_i$$

- και ας υπολογίσουμε το μέσο τετραγωνικό σφάλμα (MSE) της προσέγγισης χρησιμοποιώντας μόνο τις πρώτες m κύριες κατευθύνσεις:

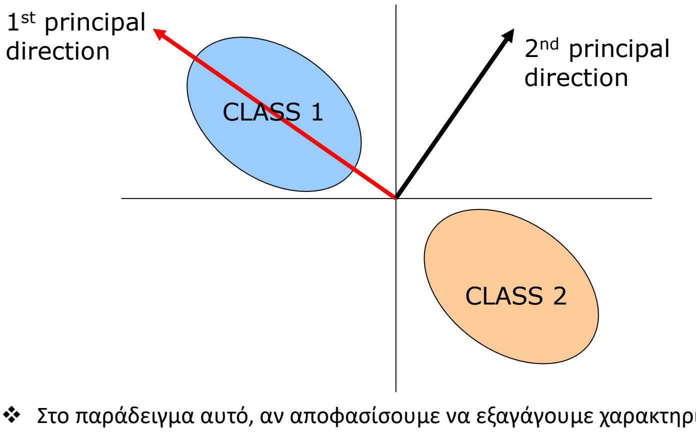
$$E\left\|\underline{x} - \underline{\hat{x}}\right\|^2 = E\left\|\sum_{i=1}^{d} y_i \hat{u}_i - \sum_{i=1}^{m} y_i \hat{u}_i\right\|^2 = E\left\|\sum_{i=m+1}^{d} y_i \hat{u}_i\right\|^2 =$$

$$= E\left(\sum_{i=m+1}^{d} \sum_{j=m+1}^{d} y_i \hat{u}_i^T y_j \hat{u}_j\right) = \sum_{i=m+1}^{d} E(y_i^2) = \sum_{i=m+1}^{d} \lambda_i$$

- Συνεπώς, η μέθοδος PCA ελαχιστοποιεί το MSE.

Principal Component Analysis (PCA)

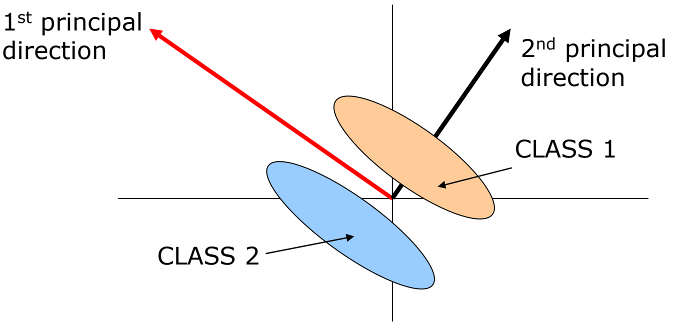
- Η μέθοδος PCA είναι μη επιβλεπόμενη. Δεν παίρνει υπόψη πληροφορία σχετικά με τις κλάσεις στις οποίες ανήκουν τα δεδομένα.



- Στο παράδειγμα αυτό, αν αποφασίσουμε να εξαγάγουμε χαρακτηριστικά μόνο στην πρώτη κύρια κατεύθυνση, το πρόβλημα θα παραμείνει διαχωρίσιμο.

Principal Component Analysis (PCA)

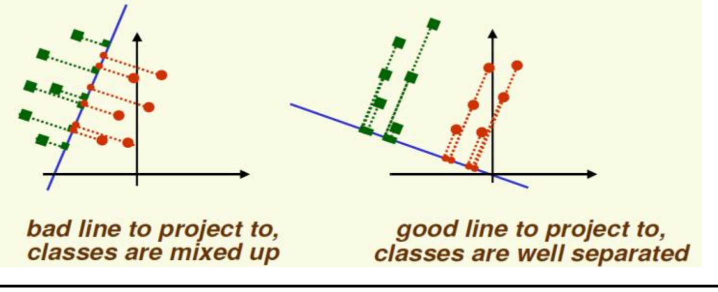
- Η μέθοδος PCA είναι μη επιβλεπόμενη. Δεν παίρνει υπόψη πληροφορία σχετικά με τις κλάσεις στις οποίες ανήκουν τα δεδομένα.



- Στο παράδειγμα αυτό, αν αποφασίσουμε να εξαγάγουμε χαρακτηριστικά μόνο στην πρώτη κύρια κατεύθυνση, μετασχηματίζουμε ένα διαχωρίσιμο πρόβλημα σε μη-διαχωρίσιμο.

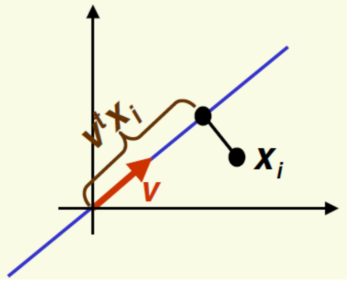
Fisher Linear Discriminant

- **Στόχος:** Να γίνει προβολή των σημείων του χώρου χαρακτηριστικών σε μία ευθεία κατάλληλης διεύθυνσης έτσι ώστε να διατηρούνται τα χρήσιμα στοιχεία του εκπαιδευτικού συνόλου για την ταξινόμηση των δεδομένων.
- Δηλαδή, θέλουμε να βρούμε την ευθεία στην οποία η προβολή των σημείων του εκπαιδευτικού συνόλου είναι όσο καλύτερα γίνεται διαχωρίσιμη
- Παράδειγμα:



Fisher Linear Discriminant

- Έστω ότι έχουμε d-διάστατα δεδομένα $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ που ανήκουν σε 2 κατηγορίες, εκ των οποίων:
  - n_1 προέρχονται από την πρώτη κατηγορία, και
  - n_2 προέρχονται από την δεύτερη κατηγορία
- Ας θεωρήσουμε τώρα προβολή των σημείων $\mathbf{x}_i$ στην ευθεία με μοναδιαίο διάνυσμα κατεύθυνσης το $\mathbf{v}$
- Το εσωτερικό γινόμενο (μονόμετρο προσημασμένο μέγεθος) $\mathbf{v}^T\mathbf{x}_i$ είναι η ορθογώνια προβολή του $\mathbf{x}_i$ πάνω στην ευθεία (δηλαδή πάνω σε έναν 1-Δ υποχώρο).



Fisher Linear Discriminant

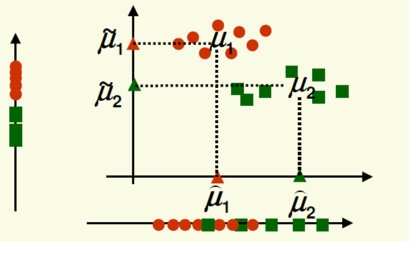
- Η προβολή του $\mathbf{x}_i$ σε ευθεία με κατεύθυνση $\mathbf{v}$ δίνεται από την $\mathbf{v}^T\mathbf{x}_i$.
- Ένας τρόπος να μετρήσουμε τον διαχωρισμό των προβολών που ανήκουν σε διαφορετικές κατηγορίες είναι με την απόσταση των μέσων τιμών των δύο κατηγοριών.
- Έστω ότι οι μέσες τιμές των προβολών των κατηγοριών 1 και 2 είναι $\tilde{\mu}_1$ και $\tilde{\mu}_2$ αντίστοιχα και έστω ότι $\boldsymbol{\mu}_1$ και $\boldsymbol{\mu}_2$ είναι οι μέσες τιμές των κατηγοριών 1 και 2. Τότε:

$$\tilde{\mu}_1 = \frac{1}{n_1} \sum_{\mathbf{x}_i \in \omega_1} \mathbf{v}^T \mathbf{x}_i = \mathbf{v}^T \frac{1}{n_1} \sum_{\mathbf{x}_i \in \omega_1} \mathbf{x}_i = \mathbf{v}^T \boldsymbol{\mu}_1$$

$$\tilde{\mu}_2 = \frac{1}{n_2} \sum_{\mathbf{x}_i \in \omega_2} \mathbf{v}^T \mathbf{x}_i = \mathbf{v}^T \frac{1}{n_2} \sum_{\mathbf{x}_i \in \omega_2} \mathbf{x}_i = \mathbf{v}^T \boldsymbol{\mu}_2$$

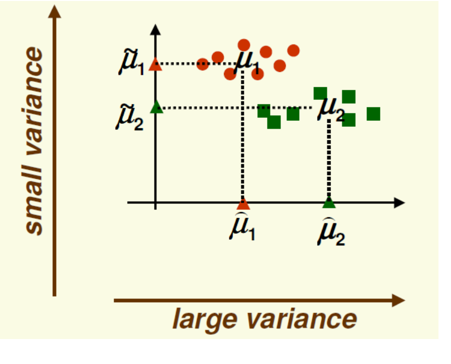
Fisher Linear Discriminant

- Η απόσταση $|\tilde{\mu}_1 - \tilde{\mu}_2|$ είναι μια καλή μετρική για το πόσο καλά διαχωρίζονται οι προβολές των δύο κατηγοριών.
- Όμως, ενώ αναμένεται όσο αυξάνει η $|\tilde{\mu}_1 - \tilde{\mu}_2|$ να βελτιώνεται και ο διαχωρισμός των κατηγοριών, αυτό δεν συμβαίνει αναγκαστικά, όπως φαίνεται στο παρακάτω παράδειγμα όπου ο διαχωρισμός στον κατακόρυφο άξονα είναι καλύτερος αν και η απόσταση των μέσων τιμών είναι μικρότερη.



Fisher Linear Discriminant

- Το πρόβλημα με την $|\tilde{\mu}_1 - \tilde{\mu}_2|$ είναι ότι δεν παίρνει υπόψη τις διασπορές των δύο κατηγοριών.



Fisher Linear Discriminant

- Έστω $y_i = \mathbf{v}^T\mathbf{x}_i$ οι προβολές των σημείων $\mathbf{x}_i$ στην ευθεία $\mathbf{v}$.
- Το scatter των προβολών της κάθε κατηγορίας είναι εξ'ορισμού:

$$\tilde{s}_1^2 = \sum_{y_i \in \omega_1} (y_i - \tilde{\mu}_1)^2$$

και

$$\tilde{s}_2^2 = \sum_{y_i \in \omega_2} (y_i - \tilde{\mu}_2)^2$$

- Δηλαδή το scatter σχετίζεται με τη διασπορά από τη σχέση $\tilde{s}_i^2 = n_i \tilde{\sigma}_i^2$

Fisher Linear Discriminant

- Η λύση του Fisher είναι η κανονικοποίηση της απόστασης $|\tilde{\mu}_1 - \tilde{\mu}_2|$ με το συνολικό scatter και η προβολή στην ευθεία $\mathbf{v}$ η οποία μεγιστοποιεί την ποσότητα $J(\mathbf{v})$:

$$J(\mathbf{v}) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

  - όπου η μεγιστοποίηση του αριθμητή αντιστοιχεί στην μεγιστοποίηση της απόστασης των μέσων τιμών των προβολών, και
  - η ελαχιστοποίηση του παρονομαστή αντιστοιχεί στην ελαχιστοποίηση της διασποράς των προβολών της κάθε κλάσης γύρω από την αντίστοιχη μέση τιμή τους
  - η λύση του Fisher δίνει την ευθεία $\mathbf{v}$ η οποία ικανοποιεί με βέλτιστο τρόπο τις δύο αυτές απαιτήσεις

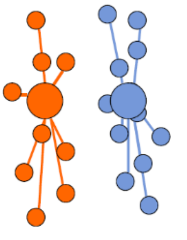
Fisher Linear Discriminant

- Αυτό που επιδιώκουμε είναι να εκφράσουμε το J(**v**) ως συνάρτηση του **v** και στη συνέχεια να το ελαχιστοποιήσουμε ως προς **v**.
- 'Εστω ότι S_1 και S_2 είναι οι πίνακες διασποράς (scatter) των δύο κλάσεων. Οι πίνακες αυτοί μετρούν τη διασπορά των αρχικών δειγμάτων στην κάθε κλάση ξεχωριστά και δίνονται από τις παρακάτω σχέσεις:

$$S_1 = \sum_{\mathbf{x}_i \in \omega_1} (\mathbf{x}_i - \mu_1)(\mathbf{x}_i - \mu_1)^T$$

και

$$S_2 = \sum_{\mathbf{x}_i \in \omega_2} (\mathbf{x}_i - \mu_2)(\mathbf{x}_i - \mu_2)^T$$



Fisher Linear Discriminant

- Ορισμός: ο πίνακας διασποράς υπολογισμένης εντός της κάθε κλάσης (***within class*** scatter matrix) συμβολίζεται με S_W και είναι το άθροισμα των πινάκων διασποράς της κάθε κλάσης:

$$S_W = S_1 + S_2$$

- Το scatter $\tilde{s}_1^2$ μπορεί τώρα να γραφεί ως εξής:

$$\tilde{s}_1^2 = \sum_{y_i \in \omega_1} (y_i - \tilde{\mu}_1)^2 = \sum_{\mathbf{x}_i \in \omega_1} (\mathbf{v}^T \mathbf{x}_i - \mathbf{v}^T \mu_1)^2$$

$$= \sum_{\mathbf{x}_i \in \omega_1} [\mathbf{v}^T (\mathbf{x}_i - \mu_1)]\,[(\mathbf{x}_i - \mu_1)^T \mathbf{v}] = \sum_{\mathbf{x}_i \in \omega_1} \mathbf{v}^T (\mathbf{x}_i - \mu_1)(\mathbf{x}_i - \mu_1)^T \mathbf{v}$$

$$= \mathbf{v}^T \Big[\sum_{\mathbf{x}_i \in \omega_1} (\mathbf{x}_i - \mu_1)(\mathbf{x}_i - \mu_1)^T \Big] \mathbf{v} = \mathbf{v}^T S_1 \mathbf{v}$$

Fisher Linear Discriminant

- Αντίστοιχα για την δεύτερη κλάση:

  $\tilde{s}_2^2 = \mathbf{v}^T S_2 \mathbf{v}$

- Συνεπώς:

  $\tilde{s}_1^2 + \tilde{s}_2^2 = \mathbf{v}^T S_1 \mathbf{v} + \mathbf{v}^T S_2 \mathbf{v} = \mathbf{v}^T S_W \mathbf{v}$

- Ας ορίσουμε και τον διαχωρισμό S_B των αρχικών μέσων τιμών:

  $S_B = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$

- Τότε

  $(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (\mathbf{v}^T \boldsymbol{\mu}_1 - \mathbf{v}^T \boldsymbol{\mu}_2)^2 = \mathbf{v}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{v} = \mathbf{v}^T S_B \mathbf{v}$

Fisher Linear Discriminant

- Συνεπώς, η J(**v**) μπορεί να γραφεί ως εξής:

$$J(\mathbf{v}) = \frac{\mathbf{v}^T S_B \mathbf{v}}{\mathbf{v}^T S_W \mathbf{v}}$$

- Η J(**v**) έχει μέγιστο για την ευθεία **v** που μηδενίζει την πρώτη παράγωγο:

$$\frac{dJ(\mathbf{v})}{d\mathbf{v}} = \frac{\left(\frac{d(\mathbf{v}^T S_B \mathbf{v})}{d\mathbf{v}}\right)\mathbf{v}^T S_W \mathbf{v} - \left(\frac{d(\mathbf{v}^T S_W \mathbf{v})}{d\mathbf{v}}\right)\mathbf{v}^T S_B \mathbf{v}}{(\mathbf{v}^T S_W \mathbf{v})^2}$$

$$= \frac{(2 S_B \mathbf{v})\,\mathbf{v}^T S_W \mathbf{v} - (2 S_W \mathbf{v})\,\mathbf{v}^T S_B \mathbf{v}}{(\mathbf{v}^T S_W \mathbf{v})^2} = \mathbf{0}$$

Fisher Linear Discriminant

- Κατά συνέπεια, ο αριθμητής πρέπει να είναι μηδέν και η ευθεία που βελτιστοποιεί τον διαχωρισμό των προβολών προκύπτει από τη λύση της εξίσωσης:

$$\mathbf{v}^T S_W \mathbf{v}\,(S_B\mathbf{v}) - \mathbf{v}^T S_B \mathbf{v}\,(S_W\mathbf{v}) = \mathbf{0} \implies$$

$$\frac{\mathbf{v}^T S_W \mathbf{v}\,(S_B\mathbf{v})}{\mathbf{v}^T S_W \mathbf{v}} - \frac{\mathbf{v}^T S_B \mathbf{v}\,(S_W\mathbf{v})}{\mathbf{v}^T S_W \mathbf{v}} = \mathbf{0} \implies$$

$$S_B\mathbf{v} - \frac{\mathbf{v}^T S_B \mathbf{v}\,(S_W\mathbf{v})}{\mathbf{v}^T S_W \mathbf{v}} = \mathbf{0} \implies$$

$$\boxed{S_B\mathbf{v} = \lambda\, S_W\mathbf{v}}$$

όπου θέσαμε $\lambda = (\mathbf{v}^T S_B \mathbf{v})/(\mathbf{v}^T S_W \mathbf{v})$.

- Η εξίσωση αυτή αποτελεί ένα γενικευμένο πρόβλημα εύρεσης ιδιοτιμών.

Fisher Linear Discriminant

- Στην περίπτωση που ο πίνακας S_W είναι πλήρους τάξης ως προς τις στήλες του (έχει τουλάχιστον από αριστερά αντίστροφο), τότε η εξίσωση μετατρέπεται στο κανονικό πρόβλημα εύρεσης ιδιοτιμών:

$$S_W^{-1} S_B \mathbf{v} = \lambda\, \mathbf{v}$$

- Όμως, γνωρίζουμε ότι το διάνυσμα $S_B\mathbf{x}$, για οποιοδήποτε $\mathbf{x}$, έχει την ίδια διεύθυνση με το $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ καθώς

$$S_B\mathbf{x} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{x} = \alpha\,(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \qquad \text{όπου } \alpha = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\mathbf{x}$$

- Συνεπώς, μπορούμε να λύσουμε άμεσα την χαρακτηριστική εξίσωση για το ιδιοδιάνυσμα

$$\boxed{\mathbf{v} = S_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}$$

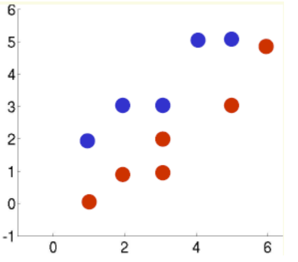
- καθώς με αντικατάσταση στην χαρακτηριστική εξίσωση έχουμε ότι:

$$S_W^{-1} S_B\,[S_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] = S_W^{-1}\,[\alpha\,(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] = \underset{\lambda}{\alpha}\,\underbrace{[S_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]}_{\mathbf{v}}$$

Fisher Linear Discriminant - Παράδειγμα

- Δεδομένα
  - Κατηγορία ω_1 με 5 δείγματα: $C_1 = \{(1,2), (2,3), (3,3), (4,5), (5,5)\}$
  - Κατηγορία ω_2 με 6 δείγματα: $C_2 = \{(1,0), (2,1), (3,1), (3,2), (5,3), (6,5)\}$
- Διευθέτηση δεδομένων σε δύο πίνακες, έναν για κάθε κατηγορία:

$$C_1 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 3 \\ 4 & 5 \\ 5 & 5 \end{pmatrix}^T \qquad C_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \\ 3 & 2 \\ 5 & 3 \\ 6 & 5 \end{pmatrix}^T$$



Ο δειγματικός χώρος

Fisher Linear Discriminant - Παράδειγμα

- Υπολογισμός μέσης τιμής κάθε κατηγορίας:

  $\boldsymbol{\mu}_1 = [3.0\ \ 3.6]^T$ και $\boldsymbol{\mu}_2 = [3.3\ \ 2.0]^T$
- Υπολογισμός πινάκων διασποράς S_1 και S_2:

$$S_1 = \begin{pmatrix} 10.0 & 8.0 \\ 8.0 & 7.2 \end{pmatrix} \quad \text{και} \quad S_2 = \begin{pmatrix} 17.3 & 16.0 \\ 16.0 & 16.0 \end{pmatrix}$$

- Ο πίνακας S_W υπολογίζεται από το άθροισμα των S_1 και S_2 :

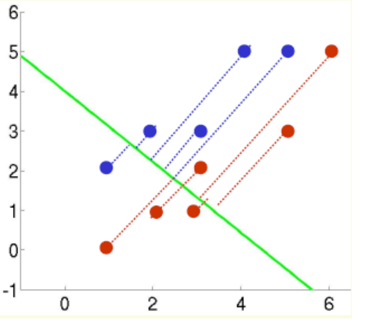
$$S_W = S_1 + S_2 = \begin{pmatrix} 27.3 & 24.0 \\ 24.0 & 23.2 \end{pmatrix} \quad \text{και επειδή } \det(S_W) \neq 0$$

$$S_W^{-1} = \begin{pmatrix} 0.39 & -0.41 \\ -0.41 & 0.47 \end{pmatrix} \quad \text{και τελικά η ευθεία για την προβολή θα είναι η}$$

$$\mathbf{v} = S_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \begin{pmatrix} 0.39 & -0.41 \\ -0.41 & 0.47 \end{pmatrix} \begin{pmatrix} -0.3 \\ 1.6 \end{pmatrix} = \begin{pmatrix} -0.78 \\ 0.88 \end{pmatrix}$$

Fisher Linear Discriminant - Παράδειγμα

- Από τη στιγμή που η ευθεία έχει τη σωστή διεύθυνση, η ακριβής θέση της δεν έχει σημασία.
- Το τελευταίο βήμα είναι ο υπολογισμός των προβολών στην ευθεία ώστε να μειώσουμε τελικά τις διαστάσεις σε μία θεωρώντας ότι η **v** έχει κανονικοποιηθεί στο μοναδιαίο διάνυσμα $[-0.65\ 0.75]^T$.



$$Y_1 = \mathbf{v}^T C_1 = [-0.65\ 0.75] \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 3 & 5 & 5 \end{pmatrix} = [0.85\ \ 0.95\ \ 0.30\ \ 1.15\ \ 0.50]$$

$$Y_2 = \mathbf{v}^T C_2 = [-0.65\ 0.75] \begin{pmatrix} 1 & 2 & 3 & 3 & 5 & 6 \\ 0 & 1 & 1 & 2 & 3 & 5 \end{pmatrix} = [-0.65\ -0.55\ -1.20\ -0.45\ -1\ -0.15]$$

Linear Discriminant Analysis (LDA)

- Η μέθοδος LDA επιλέγει τα m από τα d ιδιοδιανύσματα που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές του $S_W^{-1} S_B$.
- Έστω $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ τα ιδιοδιανύσματα του $S_W^{-1} S_B$ που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές.
- Υπολογίζουμε τον πίνακα A με τα m ιδιοδιανύσματα:

$$A = [\mathbf{u}_1 | \mathbf{u}_2 | \ldots | \mathbf{u}_m]$$

και προβάλουμε όλα τα δεδομένα στον υποχώρο που διατρέχεται από τα m ιδιοδιανύσματα:

$$\mathbf{y} = A^T \mathbf{x}$$

- Η προβολή αυτή μεγιστοποιεί την διαχωρισιμότητα των δεδομένων σε ένα χώρο μειωμένων διαστάσεων.