Lecture 12

PCA recap

1) $X \in \mathbb{R}^{d \times n}$ Project to a lower dim.

2) centralize $X_c \leftarrow X - \mu$

~~3) $\Sigma =$~~

3) Cov. $\Sigma = \frac{1}{n-1} X_c X_c^T$

4) $U \leftarrow$ eigenvectors of Σ

5) Encode, $Y = U^T X_c$

$$Y_p = U_p^T X_c \in \mathbb{R}^{p \times n}$$

$$\text{MSE}\left(U U^T x_c + \mu, X\right) \approx 0$$

$$\text{MSE}\left(U_p U_p^T x_c + \mu, X\right)$$

*) New sample $x_{new} \in \mathbb{R}^{d \times 1}$

$$y_p = U_p^T x_{new}$$

PCA

Fisher' Discriminant Analysis
Supervised dimensionality reduction

Let there be training data available for two category classification

How to project onto w while maintaining the seperability?

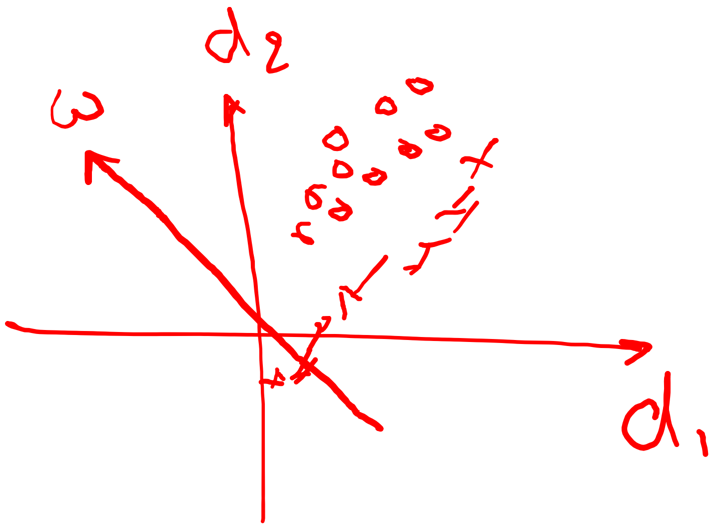
Training set:

$X_1 \in R^{d \times n_1}$, labels $t_1 \in R^{n_1 \times 1}$ →

$X_2 \in R^{d \times n_2}$, labels $t_2 \in R^{n_2 \times 1}$

$t_1 \to$ class apple

$t_2 \to$ class orange



X_1

$y_1 = \omega^T X_1 \in \mathbb{R}^{1 \times n_1}$

$\omega \in \mathbb{R}^d$

$y_2 = \omega^T X_2$

$\mu_{y_1} = \omega^T \mu_1$

$\mu_1 \rightarrow$ mean of X_1

$\mu_1 \in \mathbb{R}^{d \times 1}$

$\mu_{y_2} = \omega^T \mu_2$

$\mu_2 \rightarrow$ mean of X_2

1. we want y_1 & y_2 to be far ~~an~~ apart.

$\rightarrow$ max. the diff. b/w the means.

$w^T \mu_1 - w^T \mu_2$

distance $(w^T \mu_1, w^T \mu_2) \rightarrow$ maximized.

distance $\left(w^T x_i^1, w^T x_j^2\right) \rightarrow$ max.

reduce the cov. of each class.

Maximize the distance b/w the means.

$\max_w \left(w^T \mu_1 - w^T \mu_2\right)^2$ | min. $w^T \Sigma_1 w + w^T \Sigma_2 w$

Scatter matrix. $S_1 = \sum_i (x_i - \mu_1)(x_i - \mu_1)^T$

for class 1

$S_2 = \sum_i (x_i - \mu_2)(x_i - \mu_2)^T$ for class 2

Between class scatter S_B

FDA.

max. $w^T \overbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}^{\text{Between class scatter } S_B} w$

min $\underbrace{w^T S_1 w + w^T S_2 w}_{S_w = S_1 + S_2}$ $\rightarrow$ With class scatter matrix.

$S_w = S_1 + S_2$

$\max.\ w^T S_B w$

$\min\ w^T S_w w$

Fisher, $\max \dfrac{w^T S_B w}{w^T S_w w}$ → Rayleigh quotient

$\max. \dfrac{w^T S_B w}{w^T w}$

$\max.\ w^T S_B w$

s.t. $w^T w = 1$

$\max.\ w^T S_B w$ s.t. $w^T S_w w = 1$

$w^T S_B w - \lambda (w^T S_w w - 1)$

$S_B w = \lambda S_w w$

Generalized eigenvalue Problem.

$$S_B w = \lambda S_w w$$

$$S_w^{-1} (\mu_1 - \mu_2) \underbrace{(\mu_1 - \mu_2)^T w}_{d \times 1} = \lambda w$$

$$w \propto S_w^{-1} (\mu_1 - \mu_2)$$

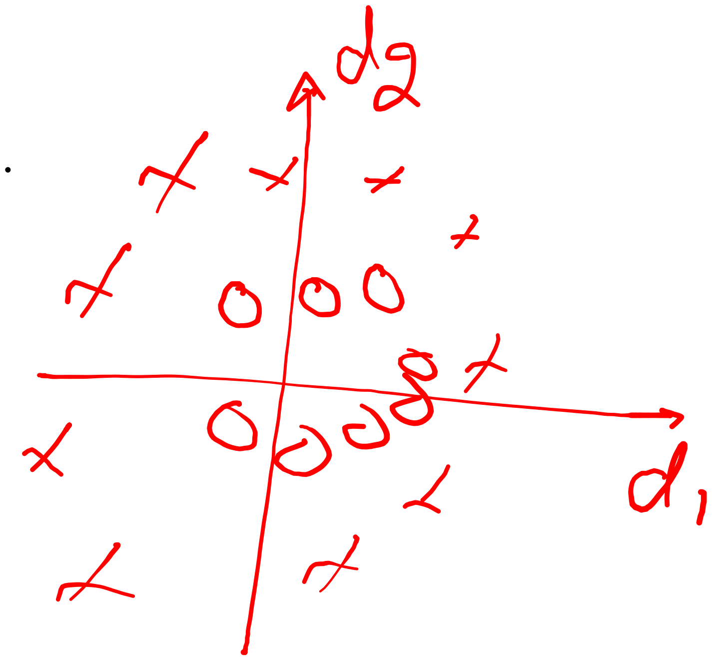
FDA vector $w = S_w^{-1} (\mu_1 - \mu_2)$

$$\max. \ \frac{w^T S w}{w^T w} \qquad w_0 \qquad w = c\,w_0$$

$c \rightarrow$ scalar

$$\max. \ w^T S w \quad s.t. \ w^T w = 1$$

$$\max. \ \frac{\cancel{c}\, w_0^T S \cancel{c}\, w_0}{\cancel{c}\, w_0^T \cancel{c}\, w_0}$$



$$y_1 = w^T X_1$$

$w \in R^{d \times 1}$, $X_1 \in R^{d \times n}$ $\qquad y_1 \in R^{1 \times n}$

Example

- Let X1, X2 be two classes. There labels are t1, t2.
- How would you apply LDA/QDA to classify a new sample xtest?
- Suppose we want to apply FDA, how do we proceed?
- Now let us apply PCA and classify?
- Now apply PCA+FDA and classify?
- Reverse FDA+PCA and classify?

μ_1, Σ_1 for X_1, μ_2, Σ_2 for X_2

$g_1(x_{test}) \leftarrow \mu_1, \Sigma_1, x_{test}$

$g_2(x_{test}) \leftarrow \mu_2, \Sigma_2, x_{test}$.

Apply FDA.

Scatter matrices. S_1 & S_2

$$S_1 \rightarrow (X_1 - \mu_1)(X_1 - \mu_1)^T$$

$$S_2 \rightarrow (X_2 - \mu_2)(X_2 - \mu_2)^T$$

Tot within class scatter $S_w = S_1 + S_2$

$$w = S_w^{-1}(\mu_1 - \mu_2)$$

Project X_1 & X_2 using w

$$y_1 = w^T X_1 \qquad y_2 = w^T X_2$$

To apply LDA.

for class 1 → μ_{y_1}, $\sigma^2_{y_1}$

for class 2 → μ_{y_2}, $\sigma^2_{y_2}$

Applying LDA/QDA to y_1, y_2

*) Multiple Discriminant Analysis.

If you have, 'c' number of classes $c > 2$

$W \in \mathbb{R}^{d \times}$

$W = [\omega_1 \quad \omega_2 \quad \cdots]$

$\omega_i \in \mathbb{R}^{d \times 1}$

$$\max \frac{|W^T S_B W|}{|W^T S_W W|}$$

$S_W = S_1 + S_2 + \cdots + S_c$

$S_B = S_T - S_W$

$S_T \rightarrow$ Total Scatter

$X_1 \in R^{d\times n_1}$, $X_2 \in R^{d\times n_2}$

$X_3 \in R^{d\times n_3}$ $\quad X_c \in R^{d\times n_c}$

$$X = [X_1, X_2 \cdots X_c] \in R^{d\times(n_1+n_2+\cdots+n_c)}$$

$$S_T = (X-\mu_x)(X-\mu_x)^T$$

$$S_B = (\mu_1-\mu_2)(\mu_1-\mu_2)^T + (\mu_1-\mu_3)(\mu_1-\mu_3)^T$$

$+$

$$X_1 = \begin{bmatrix} 1 & -1 & 0 \\ 2 & -2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} 5 & 6 & 10 \\ 0 & 1 & 11 \\ 1 & -1 & 12 \end{bmatrix}$$

$$X_3 = \begin{bmatrix} 3 & 0 \\ 2 & 0 \\ -1 & 1 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & & 5 & 6 & \cdots & 0 \\ 2 & \cdots & 0 & 1 & & 0 \\ 3 & & 1 & -1 & & 1 \end{bmatrix}$$