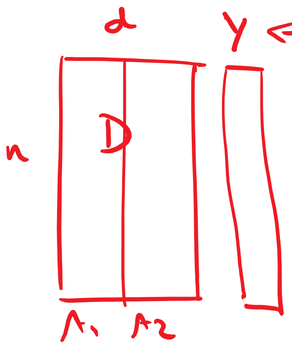
Friday, September 23, 2016 10:05 AM

Linear Discriminant Analysis (LDA)

PCA ← best variance / min error line

LDA ← find the best line



Y ← class (ground truth)

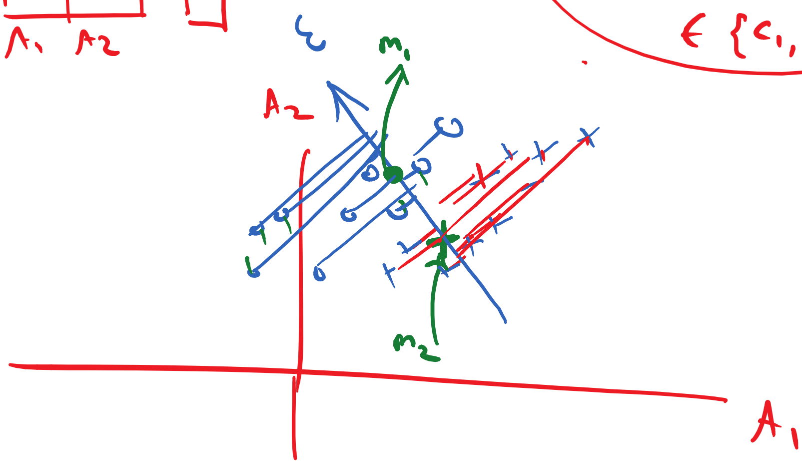
$Y \in \{+1, -1\}$

$\in \{P, N\}$

$\in \{c_1, c_2\}$

class

Categorical attribute



$D \rightarrow D_1$

$D \rightarrow D_2$

$\vec{w}$

LDA:

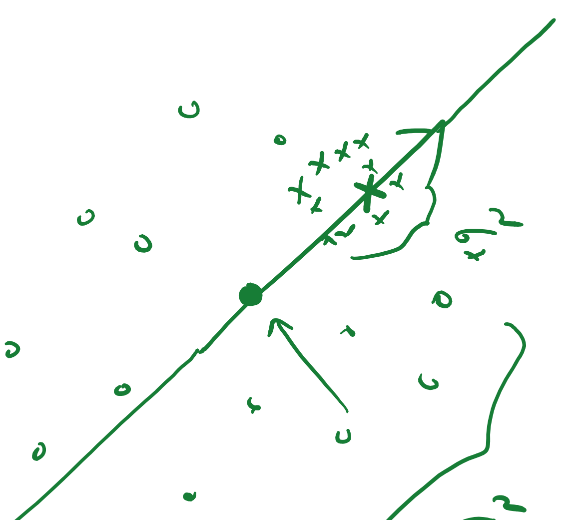
find $\vec{w}$ that separates the two classes

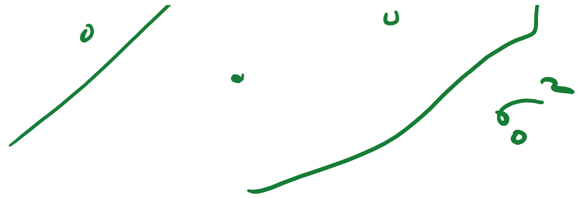
(1) means of projected points are far apart

max $(m_1 - m_2)^2$

(2) minimize the projected scatter

min $s_1^2 + s_2^2$





$\ldots S_1 + S_2$

Objective : $\max \dfrac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$

Fisher LDA

Projected Scatter : $\sum_{a_i \in C_1} (a_i - m_1)^2 = s_1^2$

$\vec{w}^T \vec{w} = 1$

$\vec{x}_i \xrightarrow{\text{Project onto } \vec{w}} a_i = w^T \vec{x}_i$

$|D_1| = n_1$

$\vec{w}^T \in \mathbb{R}^d$ $\qquad a_i \in \mathbb{R}$

$\vec{x}_i \in \mathbb{R}^d$

Then $\sigma_1^2 = \dfrac{s_1^2}{n_1}$

LDA: $\max_{\vec{w}} \dfrac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$

$\vec{\mu}_1 \leftarrow D_1$

D

$\vec{\mu}_2 \leftarrow D_2$

$$m_1 = \frac{1}{n_1} \sum_{x_i \in D_1} \vec{w}^T \vec{x}_i$$

$$= w^T \Big(\underbrace{\frac{1}{n_1} \sum_{x_i \in D_1} \vec{x}_i}_{} \Big)$$

$$= w^T \vec{\mu}_1$$

$$m_2 = w^T \vec{\mu}_2$$

$$(m_1 - m_2)^2$$

$$(w^T \vec{\mu}_1 - w^T \mu_2)^2$$

$$\left[w^T (\vec{\mu}_1 - \vec{\mu}_2) \right]^2$$

$$\underline{w^T (\vec{\mu}_1 - \vec{\mu}_2)} \; \underline{(\mu_1 - \mu_2)^T w}$$

$$(m_1 - m_2)^2 = w^T B w$$

$$w^T \left[(\vec{\mu}_1 - \vec{\mu}_2) \ldots \right]$$

$s_1^2 + s_2^2$

$s_1^2 + s_2^2$

$$s_1^2 = \sum_{x_i \in D_1} (a_i - m_1)^2$$

$$= \sum_{x_i \in D_1} \left(\underline{w}^T \vec{x}_i - \underline{w}^T \vec{\mu}_1 \right)^2$$

$$= \sum \left[w^T (\vec{x}_i - \mu_1) \right]^2$$

$$\sum_{x_i \in D_1} w^T \left[(x_i - \mu_1)(x_i - \vec{\mu}_1)^T \right] w$$

$$s_1^2 = w^T \left(\underbrace{\sum_{x_i \in D_1} (\vec{x}_i - \vec{\mu}_1)(x_i - \mu_1)^T}_{\underbrace{n_1 \cdot \Sigma_1}_{S_1}} \right) w$$

S_1 within-class scatter matrix

$$w^T \underbrace{\left[(\vec{\mu}_1 - \vec{\mu}_2)(\mu_1 - \mu_2)^T \right]}_{B} w$$

B rank-one

between class scatter matrix

$d \times d$

$$s_1^2 = w^T S_1 w$$

$$\underline{\underline{s_1^2 + s_2^2}} = w^T S_1 w + w^T S_2 w = w^T S w$$

Projected scatter

$S = S_1 + S_2$ total within class scatter

$$\boxed{\mathrm{LDA}(w): \max_{\vec{w}} \frac{w^T B w}{w^T S w}} \quad \text{s.t.} \quad w^T w = 1$$

$$\boxed{\frac{\partial LDA(w)}{\partial \vec{w}} = 0}$$

$$\boxed{B\vec{w} = \lambda S\vec{w}}$$

generalized eigen-value/vector problem.

$$\Sigma u = \lambda u$$

$$Bu = \lambda S u$$

$$S = \underbrace{n_1\Sigma_1}_{S_1} + \underbrace{n_2\Sigma_2}_{S_2}$$

symmetric PSD

S^{-1} : pseudo-inverse

$$S^{-1}B\vec{w} = S^{-1}(\lambda S\vec{w})$$

$$\boxed{(S^{-1}B)\vec{w} = \lambda\vec{w}}$$

LDA direction is the dominant eigenvector of $S^{-1}B$

how to compute $\vec{w}$ without eigenvectors

pseudo-inverse of a matrix

$$S = U\Delta U^T$$

$$= \underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{bmatrix}}_{U} \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} U^T$$

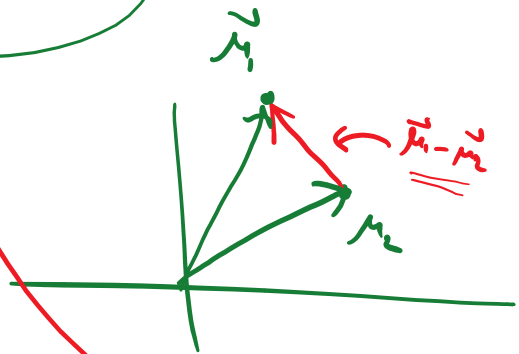
$$S^{-1} = U\Delta^{-1}U^T$$

$$= U\begin{bmatrix} 1/\lambda_1 & & & & \\ & 1/\lambda_2 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} U^T$$

$$\boxed{B\vec{w} = \lambda S \vec{w}}$$

$$B w = (\vec{\mu}_1 - \vec{\mu}_2)\underbrace{(\vec{\mu}_1 - \vec{\mu}_2)^T w}_{\alpha}$$

$$B\vec{w} = \underline{\underline{\alpha}}(\vec{\mu}_1 - \vec{\mu}_2)$$

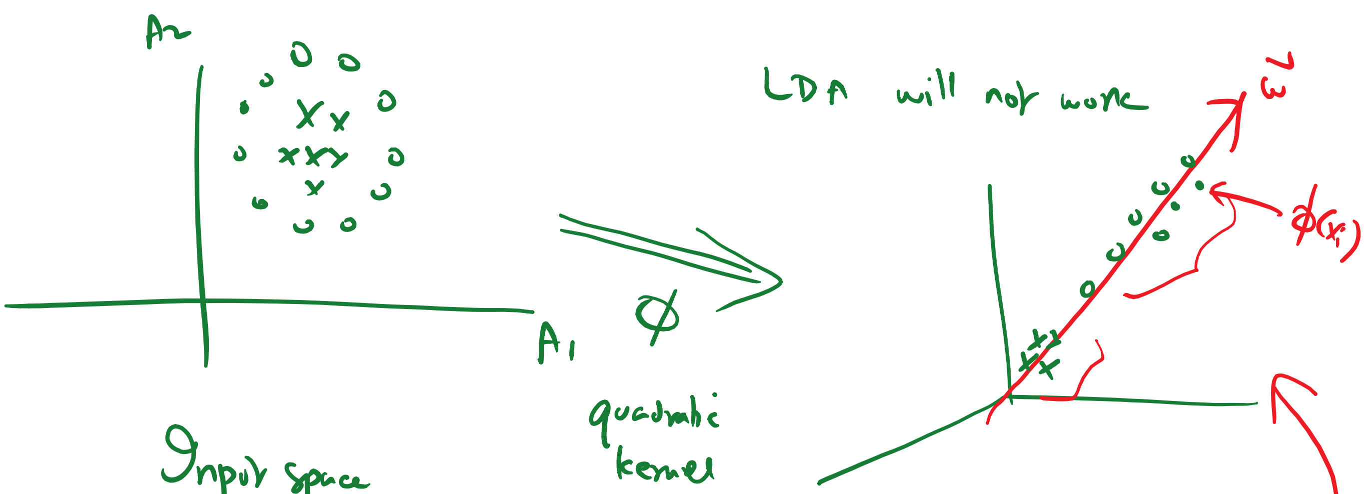


$$S^{-1}\cancel{\frac{\alpha}{\lambda}}(\vec{\mu}_1 - \vec{\mu}_2) = \vec{w}$$

$$\boxed{\vec{w} = S^{-1}(\vec{\mu}_1 - \vec{\mu}_2)} \quad \text{then} \quad \frac{\vec{w}}{\|\vec{w}\|}$$

how does S^{-1} transform $\vec{\mu}_1 - \vec{\mu}_2$

Kernel Discriminant Analysis



Input space

D

feature space

K ← kernel

$K \leftarrow$ kernel matrix

$n \times n$

Input

$$\frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

Feature space

$$\frac{(m_1^{\phi} - m_2^{\phi})^2}{(s_1^{\phi})^2 + (s_2^{\phi})^2}$$

$$\boxed{w = \sum_{i=1}^{n} c_i \phi(x_i)}$$

$\vec{w}$ is simply a linear combination of feature points

$$\phi(x_j)^T \vec{w} = \sum_{i=1}^{n} c_i \, k(x_i, x_j)$$

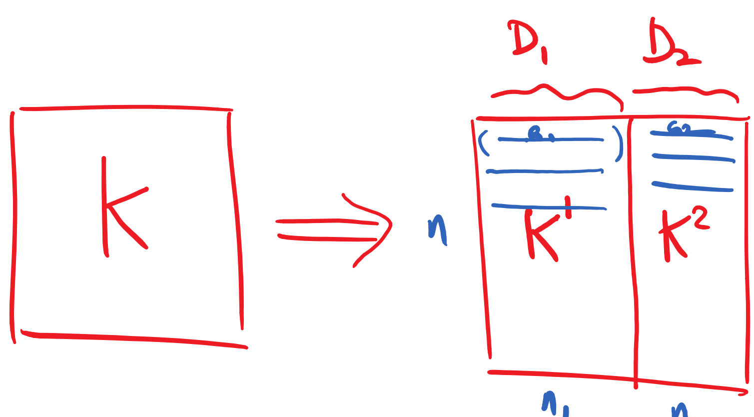
$$J(w) = \frac{w^T B w}{w^T S w}$$

$$J(\vec{w}) = \frac{w^T M w}{w^T N w}$$

M: between class scatter

N: within class scatter

$$\underbrace{N = N_1 + N_2}_{\text{from } K \text{ only}}$$



$$M = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T$$

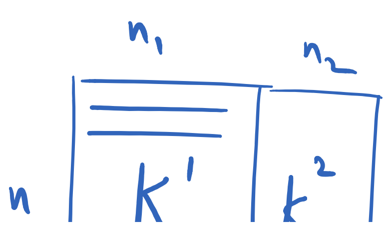
mean using K

$$\vec{m}_1 = \frac{1}{n_1}\begin{bmatrix} \text{avg for row 1 in } K^1 \\ \text{" " " 2 in } K^1 \\ \vdots \\ n \end{bmatrix}$$

$\vec{m}_1 \in \mathbb{R}^n$

$N = \underline{\underline{N_1}} + N_2$

scatter for c_1



$$n \; \begin{bmatrix} K^1 & K^2 \end{bmatrix} \; n$$

$$N_1 = \underbrace{K^1 (K^1)^T}_{\text{regular dot}} - n_1 m_1 m_1^T = \underset{n \times n_1}{(K^1)} \left(I - \frac{1}{n_1} E\right) \underset{n_1 \times n}{(K^1)^T}$$

I : identity $n_1 \times n_1$

E : matrix of all ones $n_1 \times n_1$

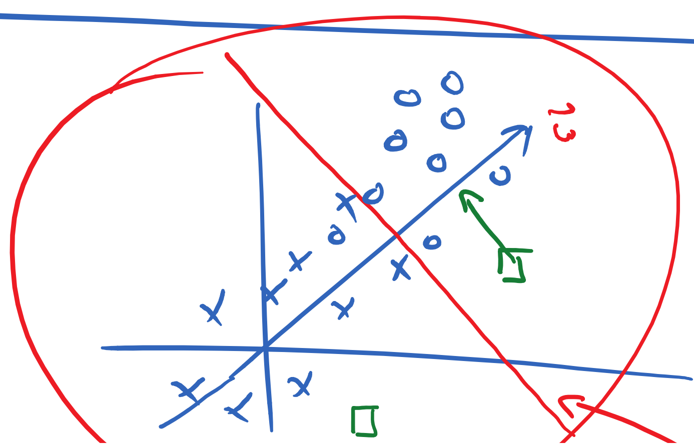
$$KDA(\vec{\omega}) : \frac{\omega^T M \omega}{\omega^T N \omega} = \omega^T \underbrace{\left(N^{-1} M\right)}\omega$$

dominant eigenvector of

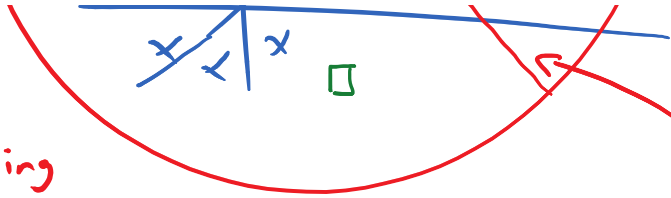
$$\boxed{\underline{\underline{N^{-1} M}} \, \omega = \lambda \omega}$$

or $Mw = \lambda N w$

generalized eigenvalue



LDA



LDA

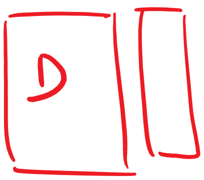
decision boundary (linear)

M

classifier

training set

y

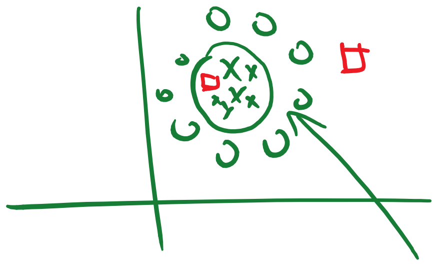


Known Classes

test point

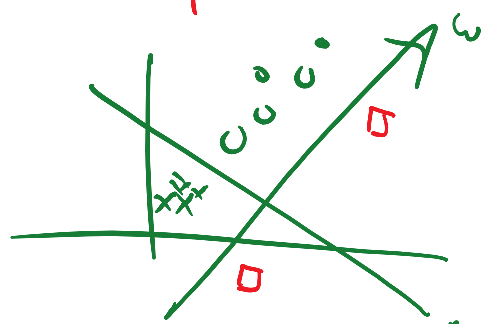
$x \rightarrow$ unknown class

$M(x) = \hat{y}$ ← predicted class



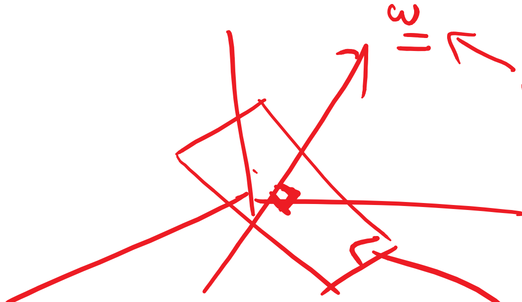
non_linear decision boundary

$\Longrightarrow$



linear

$\underline{w}$ ← normal vector



hyperplane

$$h(x) = w_0 + w_1 x_1 + w_2 x_2 + \dots w_d x_d = 0$$

$$\boxed{h(x) = w^T \vec{x} = 0}$$

all the points such that $h(\vec{x}) = 0$

all the points such that $h(\vec{x}) = 0$