Lecture 13

FDA

1. $X_1 \in$ class 1, $X_2 \in$ class 2

→ max. the distance b/w means after proj. on ω

→ min the within class scatter.

$$S_W = S_1 + S_2$$

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

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

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

$$\max_{\omega} \frac{\omega^T S_B \omega}{\omega^T S_W \omega}$$

$$\omega^T S_B \omega - \lambda(\omega^T S_W \omega - 1)$$

$$\omega = S_W^{-1}(\mu_1 - \mu_2)$$

① For multiclass.

$W \in R^{dx}$

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

$S_B = S_T - S_W$

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

$S_B w_i = \lambda_i S_W w_i$

$S_W^{-1} S_B \leftarrow$ eigenvectors

W

$W \in R^{d \times (c-1)}$

let $x_1, x_2, x_3 \rightarrow$ 3 different classes.

$y_1 = W^T x_1$

$y_2 = W^T x_2$

$y_3 = W^T x_3$

Let the data matrices be X1, X2 and X3, with classes 1, 2, 3 respectively.

Compute means.
Compute cov and scatter matrices, Sb and Sw

Now, let us project them using W. recompute means, cov and scatter matrices in terms of W and above computed Expressions.

Considering the case where you have a huge dimensionality
Issue but want to apply FDA, you may think that PCA followed
By FDA may be good for obtaining a good discriminant while
Reducing the dimensionality (complexity).

Is it good to apply them individually?

Can we do it jointly? How? Think of the objective of PCA and FDA.

① $W^T(U_p^T X_1)$

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

$w^T X \rightarrow$ PCA + FDA

We wan $w \in \mathbb{R}^{d \times 1}$

① If want to reduce dim.

let the dataset be $X_1 \in$ class1 (n samples), $X_2 \in$ class2 (n samples).

$$X = [X_1 \ X_2] \in \mathbb{R}^{d \times 2n}$$

for PCA, we want to max. var. along w

$$\max_{w} \frac{w^T (X-\mu)(X-\mu)^T w}{2n-1}, \quad \text{max. } w^T S_B w, \quad \text{min } w^T S_W w$$

$$\max_{w} \frac{w^T (X-\mu)(X-\mu)^T w}{2n-1} + w^T S_B w - \lambda (w^T S_W w - 1)$$

$$\max_x f(x) \qquad \max_x g(x)$$

$$\max_x \; f(x) + \alpha\left(g(x)\right)$$

α is a scalar.

Previously we had seen classification. Let us take some data
From property tigers -

	Sq. mtr.	Dist. citycenter (Km)	Builder reputation.	CM	y_i Price
x_1	1K	2	3.9	Kejriwal	1cr
x_2	2K	1.5	4.5	Yogi	2cr
x_3	1.5K	10	4.9	P. Vijayan	10cr
x_4	1.2K	5	3.8	Khattar	?

Regression: Given data (x_i, y_i), $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}^p$

learn or approximate $\hat{f}$ which is closer to f,

$y_i = f(x_i) + n_i$

Let us assume that data is coming from a true function f.
Goal: to find an approximate f^ which is closer to f.
Then given a test point x*, we can predict y* using f^.

.

Let us take some points

$f \rightarrow$ true function / unknown to us.

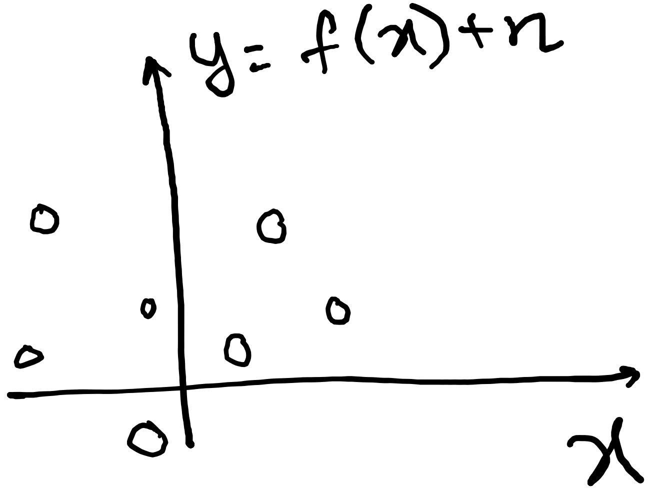
$n_i \rightarrow$ noise present in the data

$x_i \rightarrow$ data point

$y_i \rightarrow$ true label for x_i

Goal: find $\hat{f}$, find y^* for x_*, $\hat{f}(x^*)$

$\hat{f}$ to be closer to f



$x, y \in \mathbb{R}$

$\hat{f}(x) = w_1 x + w_0$

w_1, w_0 to be learned.

$\hat{y}_i = w_1 x_i + w_0$

$\hat{y}_i$ closer to y_i

reduce the error/distance.

$err_i = \phi(\hat{y}_i, y_i)$

$\phi \to$ distance/divergence

$\phi \to$ dist. function.

*) $\phi(a,b) = \phi(b,a)$

**) $\phi(a,a) = 0$

*) $\phi(a,c) \le \phi(a,b) + \phi(b,c)$

$\cancel{\phi(a,b)}$ $\phi \rightarrow$ Euclidean distance.

$\cancel{\phi}$ y_i, $\hat{y}_i$

$$\hat{y}_i = w_1 x_i + w_0$$

$$\hat{y}_i = w_2 x_i^2 + w_1 x_i + w_0$$

$$\cancel{\hat{y}_i} = X^T W = \hat{Y}$$

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 1 & & 1 \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix}$$

$$\hat{Y}_i = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\phi(y, \hat{y}) = \|y - \hat{y}\|_2^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$\|\cdot\|_2 \rightarrow l_2$ - norm. / Euclidean norm.

$$\|w\|_2^2 \rightarrow \sum_{j=1}^{d} w_j^2 \qquad w \in R^{d \times 1}$$

$$= w^T w$$

$$\|w\|_1 \rightarrow \sum_{j=1}^{d} |w_j|$$

$$\phi(y, \hat{y}) = (y - \hat{y})^T (y - \hat{y})$$

$$= y^T y - y^T \hat{y} - \hat{y}^T y + \hat{y}^T \hat{y}$$

Define X as a 2xn matrix

$W = (X'X)^{-1}XY$

Demo

$$\phi(y, \hat{y}) = y^T y - y^T X^T W - W^T X Y + W^T X X^T W$$

$$\min_W \phi(y, \hat{y})$$

$$\frac{\partial}{\partial W} \phi(y, \hat{y}) = 0$$

(x_i, y_i)

$(0, 1), \quad (1, 3) \quad (3, 2)$

$$X = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

Quad. $X = \begin{bmatrix} 0 & 1 & 9 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}$ $W = \begin{bmatrix} w_2 \\ w_1 \\ w_0 \end{bmatrix}$