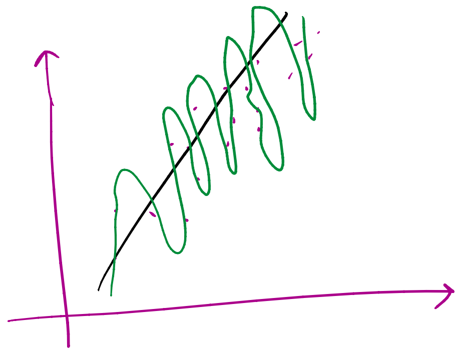
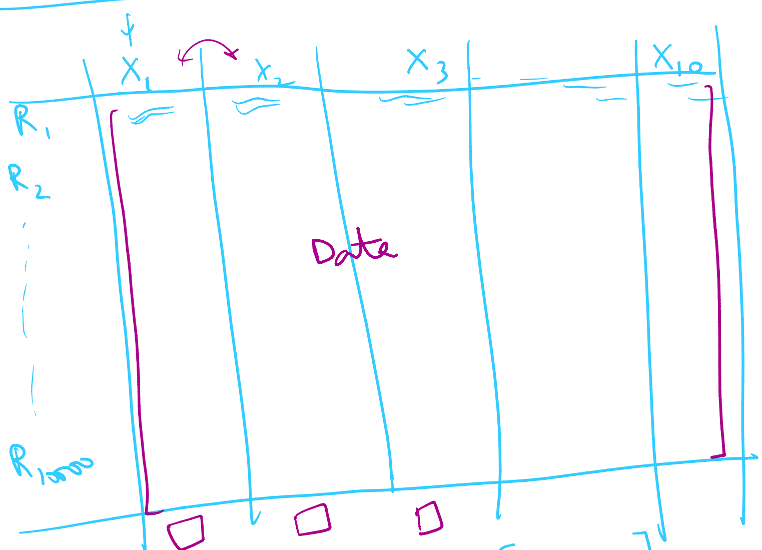
22 January 2022 13:48



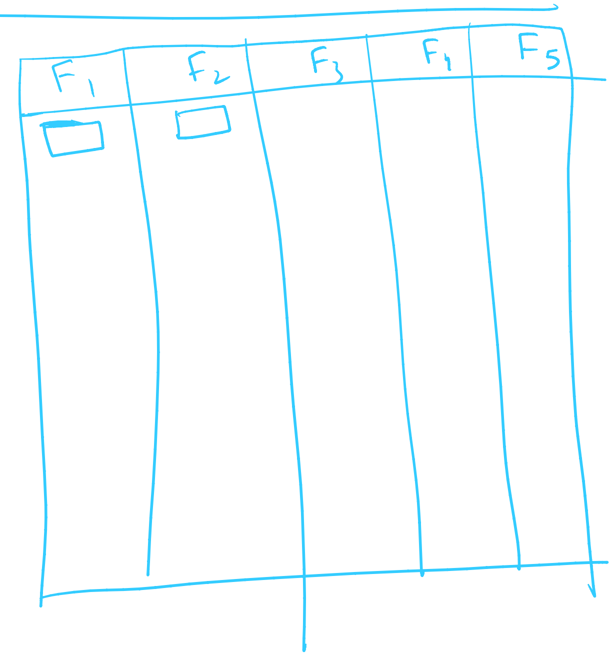
10000 data points

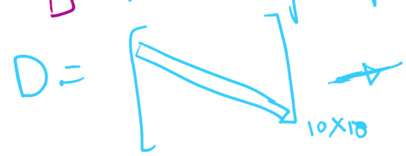
"8000 data points → Build the model

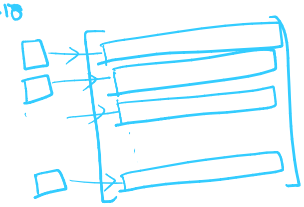
"2000 data points"



→







$\lambda_1+\lambda_2+\dots+\lambda_{10}$ = Sum of variances

$$\frac{(\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5)}{(\lambda_1+\lambda_2+\dots+\lambda_{10})}$$

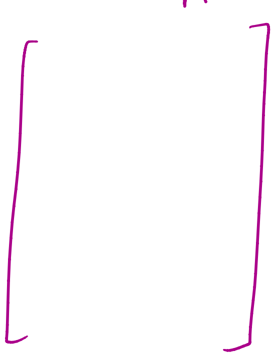
$$\frac{A\,A^T}{n}$$

~~Data~~

$\text{Data}^{\text{Trans}} \cdot \text{Data}$

10x1000 1000x10

A



① SVD – Singular Value Decomposition

② LDA → Linear Discriminant Analysis

① $\underline{\text{SVD}}$ – Singular Value Decomposition

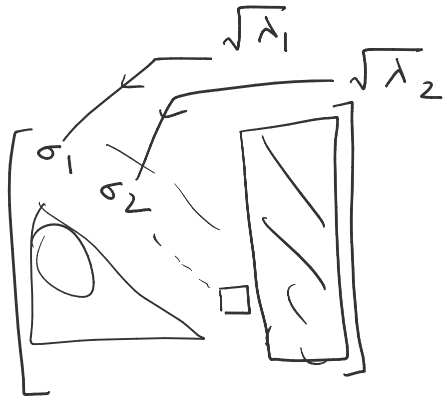
$$\underset{(m\times n)}{A} = \underset{(m\times m)}{U}\,\underset{(m\times n)}{\Sigma}\,\underset{(n\times n)}{V^T}$$

$\underset{(m\times m)}{U}$ = orthogonal eigenvectors of $\underline{A A^T}$

$\underset{(n\times n)}{V}$ = orthogonal eigenvectors of $\underline{A^T A}$

② LDA → Linear Discriminant Analysis

$\boxed{\text{'Singular Values'}}$



$$\underset{m\times n}{\begin{bmatrix} A \end{bmatrix}} = \begin{bmatrix} \; & \; \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \ddots & \sqrt{\lambda_n} \end{bmatrix} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix}$$

Concept of Rank of a Matrix

$$\begin{array}{c} v_1 \rightarrow \\ v_2 \rightarrow \end{array} \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

$2v_1 - v_2 = [0 \quad 0]$

Rank = 1

$R_i = R_i - k R_j$

$R_2 = R_2 - 2R_1$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\begin{array}{c} v_1 \\ v_2 \\ v_3 \end{array} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = [0 \quad 0 \quad 0]$

$\alpha_1 = 0 = \alpha_2 = \alpha_3$

Rank = 3

"Column Rank = Row Rank = Rank of a Matrix"

No. of non-zero eigenvalues = Rank of the matrix.

$A_{m\times n} = U \Sigma V^T$

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}\begin{bmatrix} v_1^T & v_2^T & \cdots & v_n^T\end{bmatrix}$$

$$= \underbrace{\sigma_1 \underset{(m\times 1)}{u_1} \underset{(1\times n)}{v_1^T} + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T}$$

$A = 200 \times 300$

≈ 60000

rank $(A) = \cancel{30}\ \cancel{100}\ 150$

$(m+n+1)\times r$

$501 \times \cancel{30}\ 150 = 15030$

75150

$$\begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix}_{2\times 3} \qquad \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 5 \end{bmatrix}$$

605×806

≈ 480000

50 singular values

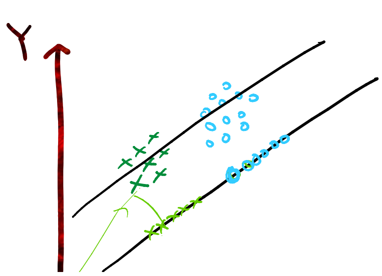
$$\underbrace{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots}_{(50)} + \sigma_r u_r v_r^T$$

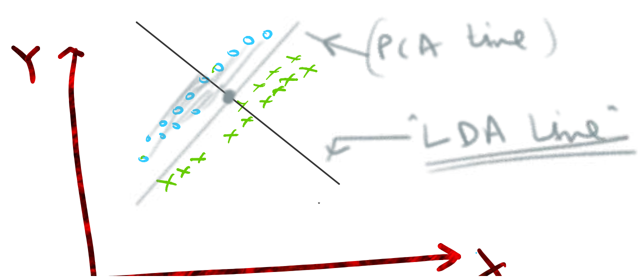
σ_1: 1, u_1: (605×1), v_1: (806×1)

$(605+806+1) \approx 1412$

$1412 \times 50 \approx 70,000$

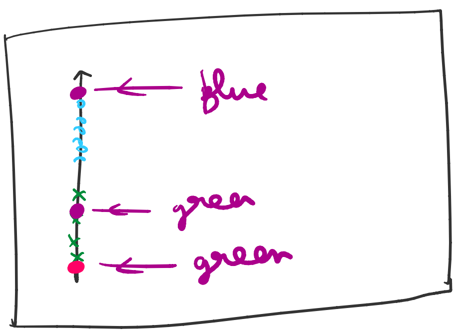
LDA

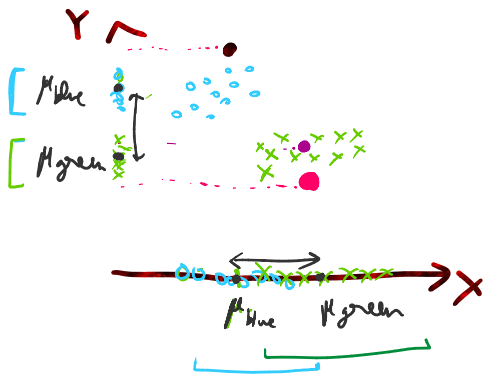




Lower the number of dimensions + preserve separability in the lower dimension.

Objective: ① Max the mean distance of the classes





N - dimension

p - dimension

② "Min the scatter"

$$S_{Blue}^2 = \sum_{blue} (y_{p_i} - \mu_{p_{Blue}})^2$$

$$S_{Green}^2 = \sum_{green} (y_{p_i} - \mu_{p_{green}})^2$$

① Max : $(\mu_{p,blue} - \mu_{p,green})^T (\mu_{p_{blue}} - \mu_{p_{green}}) = (\mu_1 - \mu_2)^2$

② Min : $S_{Blue}^2 + S_{green}^2 \Rightarrow$ Max : $\dfrac{1}{S_{Blue}^2 + S_{green}^2}$

$$\left(\text{Max: } \left[\frac{(\mu_1 - \mu_2)^2}{S_{Blue}^2 + S_{green}^2} \right] \right)$$

$$S_w = S_{Blue}^2 + S_{green}^2$$

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