(4) (a) Dimensionality Reduction Can be done by a method called Fisher's Linear Discriminant Analysis (FLDA)

↳ FLDA projects the data on the direction of maximum separation between the classes.

Assume for 2 classes class 0 and class 1

$y \in \{0, 1\}$

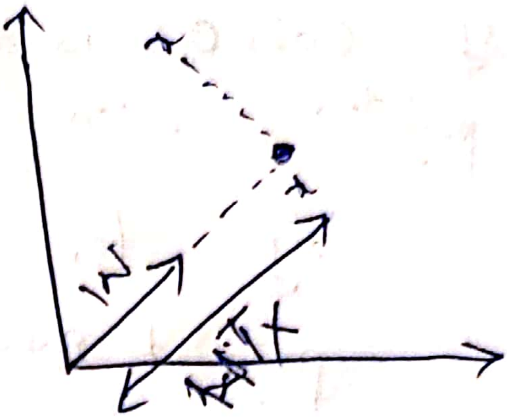
Let us assume d-dimensional data points ~~are~~

$(x_1 \cdots x_n)_{d \times 1}$

Main aim is to project all the data points on one-dimensional direction

Let w be a unit vector on which data points are projected as we ~~only~~ are considered only with direction.

$W^T x$ denotes the data points after projection.



$z = W^T x$

$x \in \mathbb{R}^d$, $W \in \mathbb{R}^d$

Considering above equation apply $W^T x$ on all data points

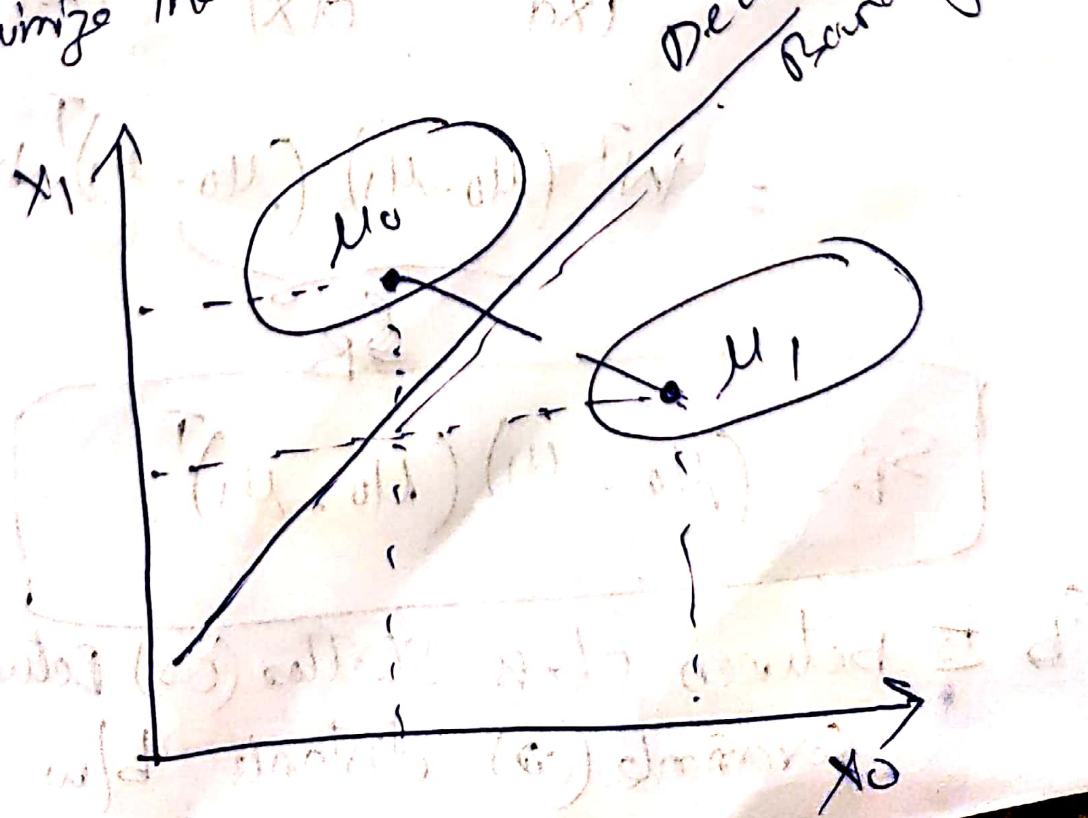
Let μ_0 is mean of class 0 and μ_1 is mean of class 1.

$\mu_0 = \frac{1}{n_0} \sum_{i:y_i=0} x_i$, $\mu_1 = \frac{1}{n_1} \sum_{i:y_i=1} x_i$

main Aim:

① to minimize variance of each class

② to maximize the distance b/w classes.



If class 0 is projected to direction of W, then projected mean for class 0 is $W^T\mu_0$.

Similarly for class 1, so

projected mean is $W^T\mu_1$

distance b/w classes is $(W^T\mu_0 - W^T\mu_1)^2$

Maximize the distance b/w classes, then

$$\max\ (W^T\mu_0 - W^T\mu_1)^2$$

Rewriting this equation

$$(W^T\mu_0 - W^T\mu_1)^2 = (W^T\mu_0 - W^T\mu_1)(W^T\mu_0 - W^T\mu_1)^T$$

$$= \underset{1\times d}{(\mu_0-\mu_1)^T}\ \underset{d\times 1}{(\mu_0-\mu_1)}\ \underset{1\times d}{W^T}\ \underset{d\times 1}{W}$$

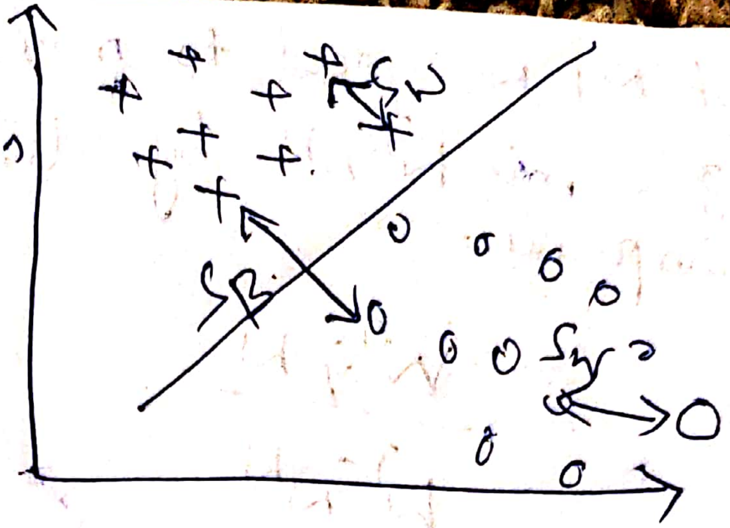
$$= W^T \underbrace{(\mu_0-\mu_1)(\mu_0-\mu_1)^T}_{S_B} W$$

$$\boxed{S_B = (\mu_0-\mu_1)(\mu_0-\mu_1)^T}$$

S_B is Between class scatter (or) Between class covariance (or) distance b/w classes.

S_W = scatter within class

= measures spread of means around the class.



Let Σ_0 be covariance of matrix of class 0
and Σ_1 be covariance matrix of class 1.

variance of projected point for class 0 (zero) and class 1 is

$$W^T \Sigma_0 W \quad \& \quad W^T \Sigma_1 W$$

~~Reo Maximize~~ minimize the variance of the classes.

$$\min_{W} W^T \Sigma_0 W + W^T \Sigma_1 W$$

$$\min_{w} W^T \underbrace{(\Sigma_0 + \Sigma_1)}_{S_W} W$$

$$\boxed{S_W = \Sigma_0 + \Sigma_1}$$

to find W, we need to max $W^T S_B W$ and minimize min $W^T S_W W$ by dividing (or) subtracting the components.

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

As per Fisher's LDA $\underset{max}{J(W)} = \frac{(\mu_0 - \mu_1)^2}{\Sigma_0^2 + \Sigma_1^2}$

$$= \frac{W^T S_B W}{W^T S_W W}$$

max $W^T S_B W$ subjective to $W^T S_W W = 1$

Apply Lagrange on the denominator to maximize

$W^T S_B W$

$$L(W, \lambda) = W^T S_B W - \lambda (W^T S_W W - 1)$$

$$\frac{dL}{dW} = 0$$

$$2 S_B W - 2 \lambda W S_W = 0$$

$$\boxed{S_B W = \lambda S_W W}$$

Scanned with CamScanner

multiply S_W^{-1} on the both sides.

S_W is a full rank matrix and inverse is possible for S_W.

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

$$\boxed{A^{-1}A = I}$$

$$\boxed{S_W^{-1} S_B W = \lambda W} \longrightarrow (1)$$

$$\boxed{S_W^{-1} S_B W = \lambda W} \Rightarrow$$ This is a eigen vector matrix

Rank of vector S_B matrix is 1 since

$$\underset{d\times d}{S_B} = \underset{d\times 1}{(\mu_0 - \mu_1)} \underset{1\times d}{(\mu_0 - \mu_1)^T}$$

$\boxed{\text{Hence the rank of matrix } S_W^{-1} S_B \text{ is } 1}$.

↳ There exists only one eigen vector. All we need to do is highest eigen vector value and this is a new projection line which is one dimension.

*** In simple terms, for 2 classes, we can rewrite equation ① as below:

$$S_W^{-1}(\mu_0-\mu_1)(\mu_0-\mu_1)^T W = \lambda W$$

Product of $(\mu_0-\mu_1)^T W$ can be considered as constant (C) since it gives a scalar matrix.

$$S_W^{-1}(\mu_0-\mu_1)\, C = \lambda W$$

$$C = (\mu_0-\mu_1)^T W$$

$$\underset{d\times d}{S_W^{-1}}\ \underset{d\times 1}{(\mu_0-\mu_1)} \propto \lambda W$$

** Product of 2 terms on LHS gives matrix $(d\times 1)$ of $d\times 1$. Hence where W has projected in direction for 2 classes we can say there exists only 1 direction.

ⓑ LDA for multiple classes.

$S_W = \sum_{k=1}^{K} S_k$ from problem ⓐ, we can say ~~that~~ for K-classes, there exists only $K-1$ directions.

Let us consider matrices for

$$W = [w_1 | w_2 \cdots w_{k-1}]_{d \times k-1}$$

Now rewriting the equation for mean and variance

$w_1^T S_B w_1,\ w_2^T S_B w_2,\ \cdots,\ w_{k-1}^T S_B w_{k-1}$

For each direction we have separate projection but we need whole variation. This can be achieved by trace which is equal to sum of covariances of all classes.

$Tr(W^T S_B W)$, $Tr(W^T S_W W)$

According to Fisher's LDA,

$$\max \frac{Tr(W^T S_B W)}{Tr(W^T S_W W)}$$

$\max Tr(W^T S_B W)$ subject $Tr(W^T S_W W) = 1$

~~Apply~~ 1

Apply Lagrange on both sides.

$$L(W,\lambda) = Tr(W^T S_B W) - \lambda(Tr(W^T S_W W) - 1)$$

$\frac{dL}{dW} = 0$

& the final equation will be by solving

$\frac{dL}{dW} = 0$,

$$\boxed{W = eig(S_W^{-1} S_B)}$$

$S_W^{-1} S_B$ has rank$(K-1)$ & has $(K-1)$ eigen vector

** S_B is sum of K different rank 1 matrices. and rank of $S_B <= K-1$ i.e. we have only $K-1$ eigen vectors and we can project data points to a subspace of dimensions at most $C-1$.

*** From all obtained eigen vectors, sort them in all descending order and select highest eigen vector value