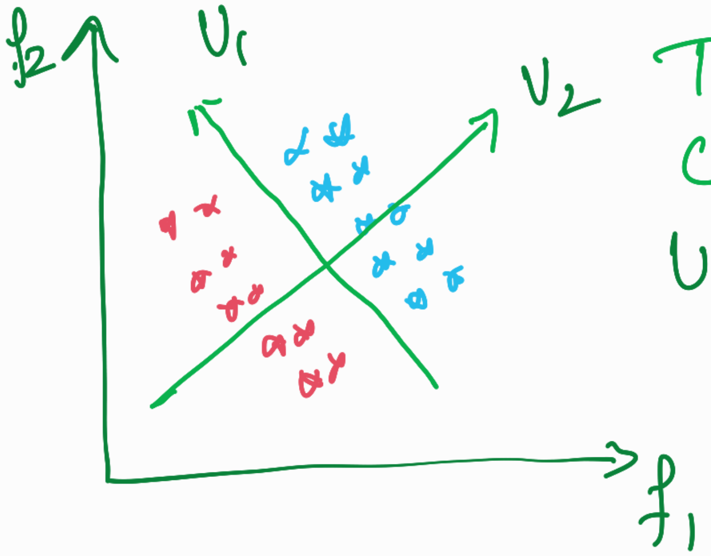
Linear Discriminant Analysis

PCA finds the best vector / direction onto which when data is projected maximum variance is preserved.

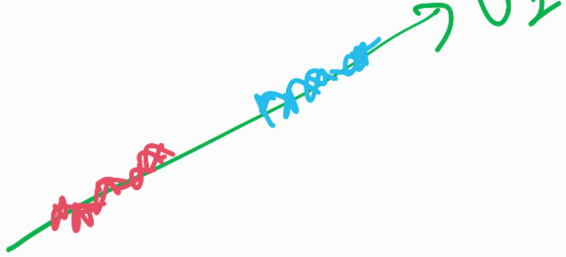
Consider the following data set



There are two classes of data,
V_1 has high Variance
V_2 has less "

When data is projected on V_1 both the classes of data get mixed and hence will not be linearly separable.

Instead of projecting on V_1 if we project on V_2, this vector V_2



seperates the classes. The classification can be performed very well in low dimensional space.

Objective : Find a direction|vector which has maximum speration of classes.

$$X = \begin{array}{c} f_1 \; f_2 \; \cdots \; f_D \\ \left[\begin{array}{ccc} & & \\ & & \\ & & \end{array} \right] \end{array}$$

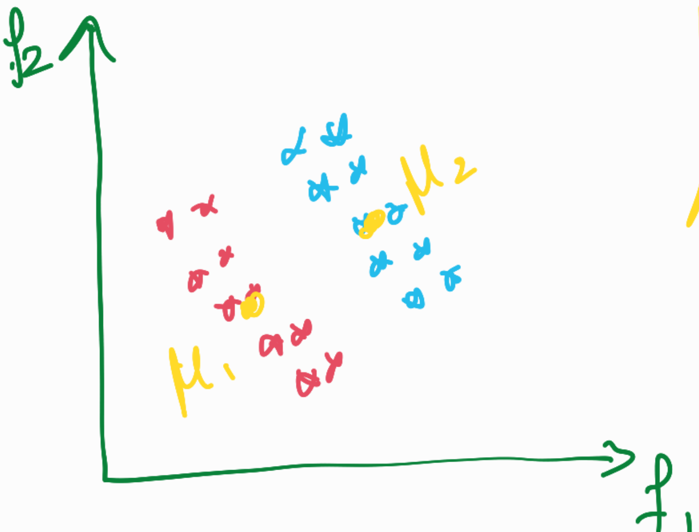
$D \rightarrow$ no of dimension

$w_i \rightarrow$ no of classes

$Y \rightarrow$ Projected space

$U \rightarrow$ required vector or direction

$$Y = U^T X$$



μ_1 is mean of class 1

μ_2 is mean of class 2

μ_i is the mean of i^{th} class

$$\mu_i = \frac{1}{N_i} \sum_{x \in \omega_1} x$$

(original space)

N_i: no of samples that belong to class ω_1

$$\tilde{\mu}_i = \frac{1}{N_i} \sum_{x \in \omega_1} U^T x$$

(transformed space)

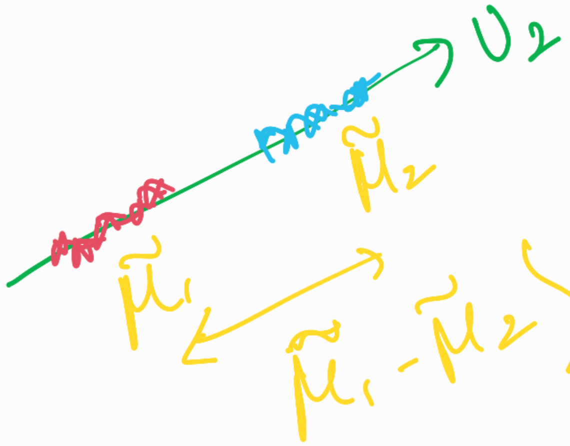
Since Σ depends on x we can move U^T outside

$$\tilde{\mu}_i = \frac{U^T}{N_i} \underbrace{\sum_{x \in \omega_1} x}_{\mu_i}$$

$$\tilde{\mu}_i = U^T \mu_i$$

$\sim$ is used for transformed space

we are looking for an axis/vector/direction which will have maximum seperation among the classes.



seperation of means needs to be maximized

Maximize $|\tilde{\mu}_1 - \tilde{\mu}_2|^2$

$$= |w^T\mu_1 - w^T\mu_2|^2$$

Find such w which maximizes $|\tilde{\mu}_1 - \tilde{\mu}_2|^2$

Using linear Algebra

$$(|A|)^2 = AA^T$$

$$|\tilde{\mu}_1 - \tilde{\mu}_2|^2 = [w^T\mu_1 - w^T\mu_2][w^T\mu_1 - w^T\mu_2]^T$$

$$(AB)^T = B^T A^T$$

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

$$= v^T S_B v$$

Where S_B is a scatter between

the classes.

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

S_B is a matrix

Suppose our data was in 2D

μ_1 is a 2D vector

μ_2 is a 2D vector

1.3	2.4	class 1	2.4	3.4	Class 2
1.6	3.4		2.3	3.0	
1.8	3.2		1.9	3.2	
2.6	2.9		2.6	3.1	
___	___	μ_1	___	___	μ_2

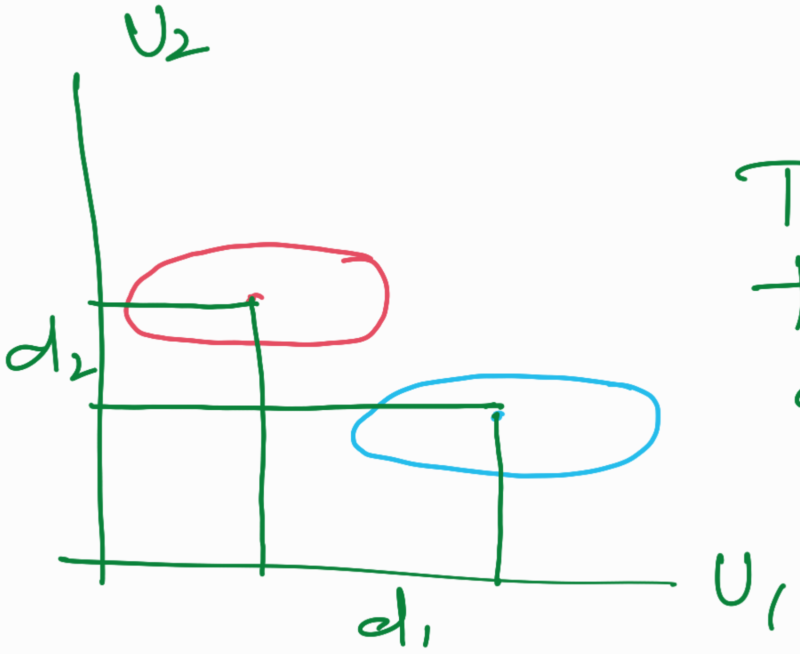
$(\mu_1 - \mu_2)_{1\times 2}$ is a 2D vector

$(\mu_1 - \mu_2)^T_{2\times 1}$ is a 2D vector

$(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ ... matrix

$(\quad)(\quad)$ 2×2 Matrix

S is the euclidean distance between the two means is a matrix form.



The distance b/w the 2 means is greater on U_1

Although U_1 has greater seperation there is a overlap between 2 classes.

Hence we need another criteria,

Scatter $$\tilde{S}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2$$

$$\tilde{S}_i^2 = \sum_{x \in \omega_i} (U^T x - U^T \mu_i)^2$$

$$= \sum_{x \in \omega_i} (U^T x - U^T \mu_i)(U^T x - U^T \mu_i)^T$$

$$= \sum_{x \in \omega_i} U^T (x - \mu_i)(x - \mu_i)^T U$$

$$= U^T \left[\sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \right] U$$

$$= U^T S_{\omega_i}^2 U$$ (we need to minimize this)

We will has 2 S_{ω_i} one for each class.

$$\tilde{S}_1^2 = U^T S_{\omega_1} U$$

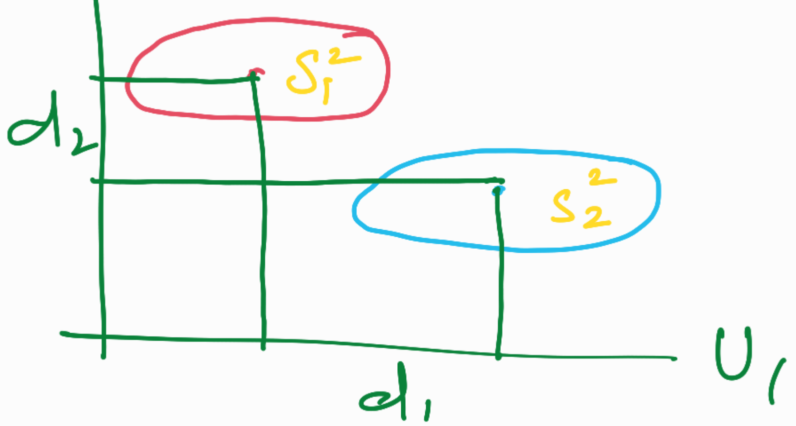
$$\tilde{S}_2^2 = U^T S_{\omega_2} U$$

$$\tilde{S}_1^2 = \tilde{S}_1^2 + \tilde{S}_2^2$$

Actual space

$$= U^T \underbrace{(S_{\omega_1} + S_{\omega_2})}_{S_\omega} U$$

U_2



Cost function

max seperation between means $U^T S_B U$

minimize scatter $U^T S_W U$

The cost function cannot maximize one & minimize the other, hence we define a function which maximize but still takes care of both

$$J = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{S}_1^2 + \tilde{S}_2^2} = \frac{U^T S_B U}{U^T S_W U}$$

If $U^T S_W U$ is min cost will be max

$U^T S_B U$ is max cost will be maximum

To maximize the cost

U If $y = \frac{u}{v}$ where u & v are differentiable

$$\frac{dy}{dx} = \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\frac{d}{dx}(u) - u\frac{d}{dx}(v)}{v^2}$$

$$\frac{\partial J}{\partial U} = \frac{\partial}{\partial U}\left(\frac{U^T S_B U}{U^T S_W U}\right)$$

$$= \frac{U^T S_W U \frac{\partial U^T S_B U}{\partial U} - U^T S_B U \frac{\partial U^T S_W U}{\partial U}}{\left(U^T S_W U\right)^2}$$

$$\frac{\partial (U^T S_B U)}{\partial U} = 2 S_B U$$

$$\frac{\partial (U^T S_W U)}{\partial U} = 2 S_W U$$

$$= \frac{U^T S_W U \times 2 S_B U - U^T S_B U \, 2 S_W U}{(U^T S_W U)^2}$$

$$0 = U^T S_W U \cdot S_B U - U^T S_B U \cdot S_W U$$

Divide by $\underbrace{U^T}_{1\times d}\underbrace{S_W}_{d\times d}\underbrace{U}_{d\times 1}$ = scalar

$$= S_B U - \left(\frac{U^T S_B U}{U^T S_W U}\right) S_W U$$

If $\lambda = \dfrac{U^T S_B U}{U^T S_W U}$

$$= S_B U - \lambda S_W U$$

$$S_B U = \lambda S_W U$$

Multiply S_W^{-1} both sides

$$S_W^{-1} S_B U = \lambda \underbrace{S_W^{-1} S_W}_{\text{Identity}} U$$

$$S_W^{-1} S_B U = \lambda U$$

which is of the form

$$Ax = \lambda x$$

v is the eigen vector of

$(S_W^{-1} S_B)$ matrix

The eigen vector corresponding to max λ will give you the direction of max seperation.

We can choose top k dimensions