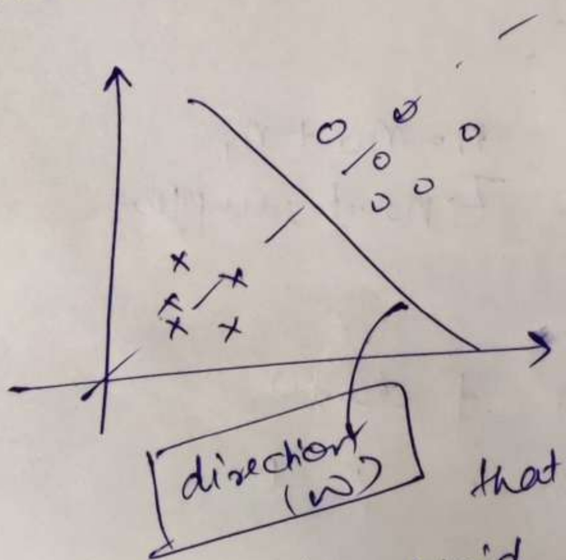
Mathematical Expression of LDA

Provides max. class separability.



We have to project data space into one particular dimension, we can project it to x-axis or y-axis but when we do that, all the points get mixed up. So, to avoid this, we need to find direction (w) that provides max. separability.

If we use dotted line, then data mixed up. So, if we use solid line which can be found by some classifier that gives max. separable boundary.

How do you find that direction (w) is good? That can be seen by looking at the data points. Which is done by;

S_W → within class scatter matrix

S_B → inter class " "

So, we try to achieve solid line (hyperplane) as in SVM

$$\underline{\underline{W^T x + b}}$$

$\{x_i, y_i\}$ $\quad y_i \in \{0, 1\}$ $\quad n = n_0 + n_1$

$x_i \rightarrow$ No. of features

$y_i \rightarrow$ class

$n \rightarrow$ no. of samples

Class 0 $\Rightarrow y_i = 0$ $\quad M_0 \rightarrow$ Mean of class 0

Class 1 $\Rightarrow y_i = 1$ $\quad M_1 \rightarrow$ Mean " " 1

$$M_0 = \frac{1}{n_0} \sum_{x_i \in C_0} x_i$$

$$M_1 = \frac{1}{n_1} \sum_{x_i \in C_1} x_i$$

We need to project it into higher dimension having means m_0 & m_1

mean $W^T x_i = z_i$

$$m_0 = W^T M_0$$

$$m_1 = W^T M_1$$

So, cost function –

$$J(w) = m_0 - m_1$$

max.

We try to separate the differences b/w means. But this would produce less effect on separability. So, we will take variance as it would give much greater effect.

σ_0^2 = variance of class 0

σ_1^2 = " " " 1

$$\sigma_0^2 = \sum_{x_i \in C_0} (w^T x_i - m_0)^2, \quad \sigma_1^2 = \sum_{x_i \in C_1} (w^T x_i - m_1)^2$$

$$\boxed{\underset{\text{max.}}{J(w)} = \frac{(m_0 - m_1)^2}{\sigma_0^2 + \sigma_1^2}}$$

→ max. mean deviation relative to sample variances.

$$(m_0 - m_1)^2 = (w^T M_0 - w^T M_1)^2$$

$$= w^T [(M_0 - M_1)(M_0 - M_1)^T] w$$

$$= w^T S_B w$$

→ separation of class means i.e. b/w class scatter matrix

$$\sigma_0^2 = \sum_{x_i \in C_0} (w^T x_i - m_0)^2$$

$$= \sum_{x_i \in C_0} (w^T x_i - w^T M_0)^2$$

$$= W_T \left[\underbrace{\sum_{x_i \in C_0} (x_i - M_0)(x_i - M_0)^T}_{} \right] w$$

$$= W_T \; \underbrace{S_W}_{\text{within class scatter Matrix.}} \; w$$

within class scatter Matrix.

$$\therefore \boxed{\underset{max}{J(w)} = \frac{w^T S_B w}{w^T S_W w}}$$

Relationship b/w S_B & S_W:

$$\frac{d\,J(w)}{d\,w} = \frac{w^T S_W w (2 S_B w) - w^T S_B w (2 S_W w)}{(w^T S_W w)^2}$$

$$= \frac{2 S_B w}{\underbrace{w^T S_W w}_{scalar}} - \frac{\overbrace{w^T S_B w}^{Scalar} (2 S_W w)}{\underbrace{(w^T S_W w)^2}_{scalar}} = 0$$

$$S_W w = \underbrace{\lambda}_{\text{Scalar Parameter.}} S_B w$$

Scalar Parameter.

Need to represent S_W in invertible form to become equations finite.

$$\sigma_0^2 + \sigma_1^2 = W^T S_W W$$

$$S_W = \sum_{x_i \in C_0} (x_i - M_0)(x_i - M_0)^T +$$

$$\sum_{x_i \in C_1} (x_i - M_1)(x_i - M_1)^T$$

This is Rank 1 effecient as represented in two different forms.

This is efficient in covariance.

Now, $S_W W = \lambda S_B W$

$\therefore W = S_W^{-1} S_B W$

$$\boxed{W = S_W^{-1} (M_0 - M_1)(M_0 - M_1)^T W}$$

$$\boxed{W_{max} = S_W^{-1} (M_0 - M_1)}$$

$\underline{W^T x} + b$ → using ROC analysis → use different values for distribution and where achieve max. class separability. use that 'b'.

$W^T x$ → we calculated this