9/4/2019 [new]

LDA: Linear Fisher Discriminator. - Derivation

LDA is a special / specific case of MDA.

- only two classes
- and the approach to separate two classes is called Fisher Discriminator also called as Linear Fisher Discriminator.

Fisher Discriminator:

Given:

n d-dimensional feature vectors

$$
\left.\begin{aligned}
x_1 &= (x_{11}, \dots x_{1d}) \\
x_2 &= (x_{21}, \dots x_{2d}) \\
&\vdots \\
x_n &= (x_{n1}, \dots x_{nd})
\end{aligned}\right\}
$$

out of these
n_1 no. of F.V $\in \omega_1$
n_2 no. of F.V $\in \omega_2$

n_1 = no. of F.V from class ω_1

n_2 = no. of F.V from class ω_2.

w = direction of projection.

$\| w \| = 1$ (unit vector in the direction of projection line)

$\omega_1 = (x_1, x_2 \dots, x_{n_1}) = n_1$ no of samples

$\omega_2 = (x_1, x_2 \dots, x_{n_2}) = n_2$ no. of samples

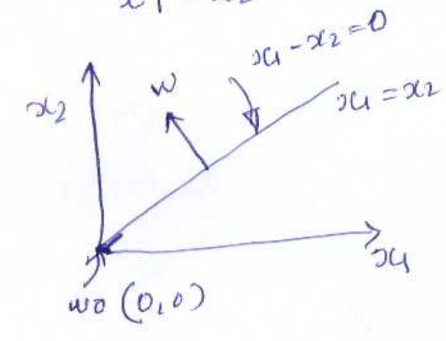
Suppose If a take orthogonal projection on data x_i and I get y_i

$$\boxed{y_i = w^t x_i}$$
$$\Downarrow$$
$$\boxed{y_i = w^t x_i}$$

$$[w_1 \ w_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$w_1 x_1 + w_2 x_2 = 0$ and this passes through origin.

$$x_1 = x_2$$
$$x_1 - x_2 = 0$$
$$w_1 x_1 - w_2 x_2 = 0$$
$$1 \cdot x_1 - 1\, x_2 = 0$$
$$\underset{w^t}{(1 \ \ -1)} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$
$$x_1 - x_2 = 0$$



If I take projections on d-dimensional F.V $x_1, \ldots, x_n$

$y_1, y_2, \ldots, y_n$ [these are projected vectors].

$$\underbrace{y_1, y_2, \ldots}_{n_1 \in \omega_1} \ \underbrace{\ldots, y_n}_{n_2 \in \omega_2}$$

let m_1 & m_2 is the mean of class ω_1 $\forall x \in \omega_1$ and class ω_2 $\forall x \in \omega_2$ ②

$$\boxed{m_1 = \frac{1}{n_1} \sum_{\forall x \in \omega_1} x}$$

$\left[\text{set } x = \underbrace{(x_1, x_2 \ldots x_{n_1})}_{n_1 \text{ no. of samples.}} \right]$ and

$$\boxed{m_2 = \frac{1}{n_2} \sum_{\forall x \in \omega_2} x}$$

set $x = \{x_1, x_2 \ldots x_{n_2}\}$
n_2. no. of samples

$\omega_1 = (x_1, x_2, \ldots x_{n_1})$

$\omega_2 = (x_1, x_2, \ldots x_{n_2})$

when I compute the projection of the mean vector, the projected mean vector m_1 can be written as $\tilde{m}_1$.

$$\tilde{m}_1 = \frac{1}{n_1} \sum_{\forall y \in \omega_1} y$$

← projected vector of x.
$y = w^t x$.

$$\tilde{m}_1 = \frac{1}{n_1} \sum_{\forall x \in \omega_1} w^t x$$

$$= \frac{1}{n_1} w^t \sum_{\forall x \in \omega_1} x.$$

$$= w^t \frac{1}{n_1} \sum_{\forall x \in \omega_1} x = w^t m_1$$

$$\boxed{\tilde{m}_1 = w^t m_1}$$

$\tilde{m}_1 \Rightarrow$ projection of mean of vectors of ~~set~~ class ω_1 is nothing but $w^t m_1$.

In the same mannar, I can compute, the projection of mean of samples of class ω_2 which is nothing but

$$\boxed{\tilde{m}_2 = w^t m_2}$$

Now, what is the Distance between these two projection mean?

<u>Distance between projected means:</u>

$$|\tilde{m}_1 - \tilde{m}_2| = |w^t m_1 - w^t m_2|$$

$$= |\underbrace{w}_{\text{vector } w}{}^t (m_1 - m_2)|$$

from, this I can ~~simply~~ increase the distance between $|\tilde{m}_1 - \tilde{m}_2|$ by simply ~~scaling~~ scaling of '<u>w</u>'.

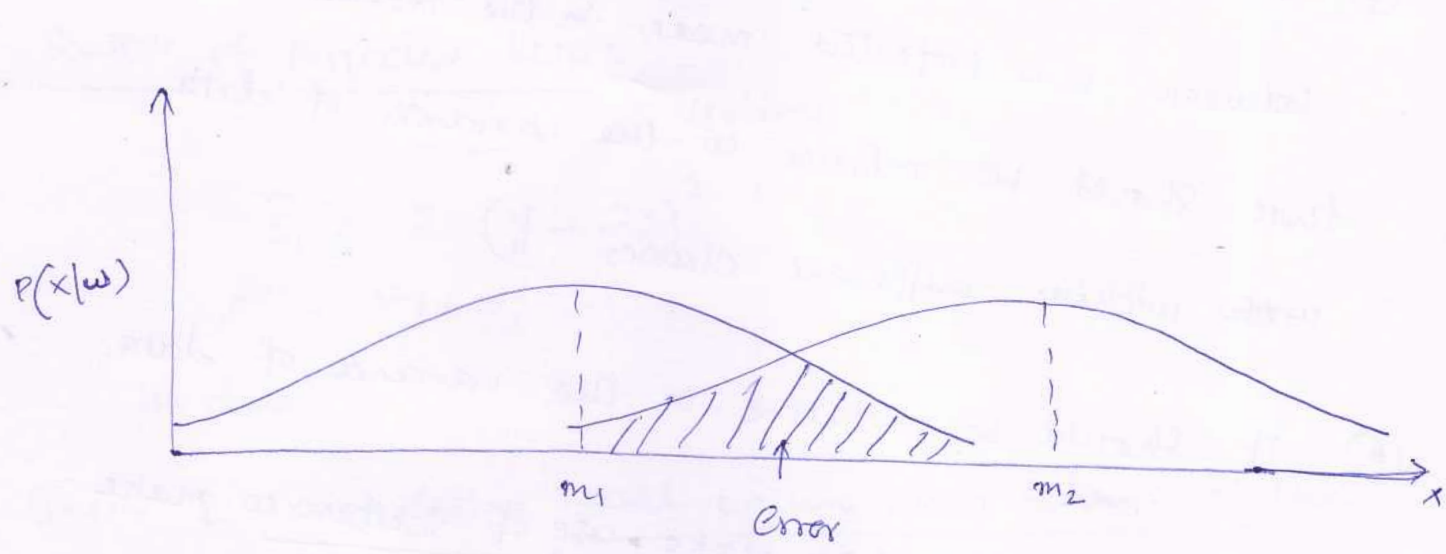
Here I assume $\|w\| = 1$. If the $\|w\| > 1$ (more than 1) then the distance between $|\tilde{m}_1 - \tilde{m}_2|$ will go on increasing.

$\boxed{\text{* This ensures that I can increase the separability between two classes}}$

But, the question is how much should I increase to separate the two classes?

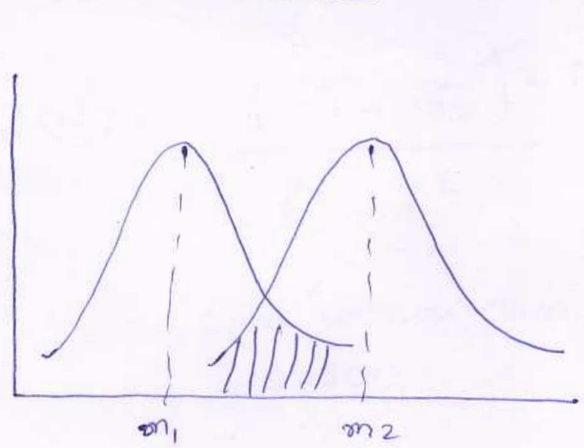
Illustration:

If the variance is large:



We should have the difference between m_1 and m_2 should be quite large so that the error of classification is reduced.

If the variance is small:



* In this case, we don't need that much separation which we need in the above case.

* The separation between m_1 and m_2 is much less than the above ~~case~~ case. (even then the error is minimized)

* So how much should be the difference between two projected mean in the reduced space, that should be relative to the variance of data ~~with~~ within different classes.

* It should be relative to the variance of data.

* Accordingly, we can make use of scatter to make some criterion function.

Scatter of projected data:

$$\tilde{s}_i^{\,2} = \overbrace{\sum_{\forall y \in \omega_i} (y - \tilde{m}_i)^2}^{\text{variance}} \Rightarrow \text{we have not normalized } \frac{1}{n}.$$

ith class

Fisher Discrimant method:

Scatter of projected data:

$$\tilde{S}_i = \sum_{\forall y \in \omega_i} \overbrace{\left(y - \tilde{m}_i\right)^2}^{\text{Variance.}}$$

ith class

Then, I can define total within class scatter of the projected samples.

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

we want the distance between two mean should be relative to this total scatter 'S'. So we can define the criterion function as follows.

difference b/w projected mean

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\underbrace{\tilde{s}_1^2 + \tilde{s}_2^2}_{\text{total within class scatter.}}} \Rightarrow \text{as large as possible w.r.t the total within class scatter.}$$

we want to maximize $J(w)$ which is the ratio of $|\tilde{m}_1 - \tilde{m}_2|^2$ upon $\tilde{s}_1^2 + \tilde{s}_2^2$. $J(w)$ is ratio of inter class scatter to intra class scatter.

$$J(w) = \frac{\text{Inter class}}{\text{Intra class}} = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Fisher Discriminant maximize the

$$J(w) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

The value of 'w' which maximizes J(w) is the projection direction.

<u>J(.) in terms of 'w'</u>

S_i $S_w \rightarrow$ total within class scatter.

Scatter within ith class

$$S_i = \sum_{\forall x \in \omega_i} (x - m_i)(x - m_i)^t$$

Scatter for individual class (ith class)

we can define, total within class scatter as

$$S_w = \sum_{\forall i} S_i$$ here, we consider only two classes

$$\boxed{S_w = S_1 + S_2}$$

Scatter of projected samples:

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

$$\tilde{S}_i^2 = \sum_{\forall x \in \omega_i} (w^t x_i - w^t m_i)^2 \qquad y_i = w^t x_i$$

by rearranging this,

$$\cancel{\tilde{S}_i^2 = \sum_{\forall x \in \omega_i} w^t (x_i - m_i)(x_i - m_i)^t w}$$

$$\tilde{S}_i^2 = \sum_{\forall x \in \omega_i} (w^t x - w^t m_i)^2$$

$$= \sum_{\forall x \in \omega_i} (w^t x - w^t m_i) \cdot (w^t x - w^t m_i)$$

$$= \sum w^t (x - m_i) \cdot \underbrace{w^t (x - m_i)}$$

$$= \sum w^t (x - m_i) \cdot (x - m_i)^t w.$$

$$= w^t \left[\sum_{\forall x \in \omega_i} (x - m_i)(x - m_i)^t \right] \cdot w$$

$$\boxed{\tilde{S}_i^2 = w^t S_i w.}$$

then the sum of $\tilde{s}_1^2 + \tilde{s}_2^2$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^t S_1 w + w^t S_2 w$$

$$= w^t (S_1 + S_2) w$$

$$= w^t \underbrace{S_w}_{\text{total within class scatter}} w.$$

total within class scatter [refer page (4)]

In the same mannar,

Separation of the projected means.

$$(\tilde{m}_1 - \tilde{m}_2)^2 = (w^t m_1 - w^t m_2)^2$$

$$= (w^t m_1 - w^t m_2) \cdot (w^t m_1 - w^t m_2)$$

$$= w^t (m_1 - m_2) \cdot \underbrace{w^t (m_1 - m_2)}$$

$$= w^t \underbrace{(m_1 - m_2)}_{\text{row vector}} \cdot \underbrace{(m_1 - m_2)^t}_{\text{coloum vector}} w$$

$$= w^t \underbrace{(m_1 - m_2)(m_1 - m_2)^t}_{\text{outer product of '2' vectors}} w$$

it rank is atmost 1.

$$= w^t \underbrace{S_B}_{\text{between class scatter}} w \Rightarrow$$ is a vector in the direction of $(m_1 - m_2)$

between class scatter

> **Observation:** $w^t \underbrace{S_B w}$ $\Rightarrow$ is a vector in the direction of $(m_1 - m_2)$

In particular, for any w, $S_B w$ is in the direction of $m_1 - m_2$, and S_B is quite singular.

$$J(\omega) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

in terms of 'w'

between class scatter

$$J(\omega) = \frac{w^t S_B w}{w^t S_W w} \Rightarrow \text{generalized Rayleigh coefficient.}$$

total within class scatter

maximize ~~this~~ ratio by varying vector 'w' and 'w' gives the projection direction.

Take a derivative w.r.t 'w' and equate it to zero.

$$J(\omega) = \frac{w^t S_B w}{w^t S_W w} = \frac{u}{v} = \frac{v \cdot u' - u \cdot v'}{v^2}$$

$$\frac{d}{dw} J(w) = \frac{w^t S_W w \cdot \frac{d}{dw}(w^t S_B w) - w^t S_B w \frac{d}{dw}(w^t S_W w)}{(w^t S_W w)^2}$$

$$0 = \frac{w^t S_W w\, [2 S_B w] - w^t S_B w\, [2 S_W w]}{\cancel{(w^t S_W w)^2}}$$

$$w^t S_W w \left[2 S_B w\right] = w^t S_B w \left[2 S_W w\right]$$

Divide $w^t S_W w$

$$2 S_B w = \frac{w^t S_B w}{w^t S_W w}\, 2 S_W w.$$

$$\cancel{2} S_B w = \lambda \cancel{2} S_W w$$

$$\boxed{S_B w = \lambda S_W w}$$ ⇒ generalized eigenvalue problem

If S_W is non-singular then S_W^{-1} exists.

Singular matrix = If the determinant of matrix $|A| = 0$ then singular

$A^{-1} = \frac{1}{|A|} [\ \ ]$

Inverse does not exist for singular matrix

$$\boxed{S_W^{-1} \underbrace{S_B w}_{(m_1 - m_2)} = \lambda w}$$ ⇒ eigenvalue problem

eigen vector of $S_W^{-1} S_B$ and λ is the corresponding eigen value.

It is unnecessary to solve for the eigen values & vectors of $S_W^{-1} S_B$ due to the fact that $S_B w$ is always in the direction of $m_1 - m_2$

The scaling factor of 'w' (in λw) is not important [in λw]

[$\cancel{S_B w = (m_1 - m_2)}$] but the direction of '$w$' is so important.

$$S_W^{-1} \underbrace{S_B w} = \lambda w$$

$$S_W^{-1} K (m_1 - m_2) = \lambda w$$

$$S_W^{-1} \left(\frac{K}{\lambda}\right) (m_1 - m_2) = w$$

$$S_W^{-1} \cancel{(K_1)} (m_1 - m_2) = w$$

$$\boxed{w = S_W^{-1} (m_1 - m_2)}$$