Linear Discriminanat Analysis

Dr. Pritam Anand.
Assistant Professor,
DA-IICT, Gandhinagar.

Projection

- The projection of any n dimensional point x_i onto the vector w is given as

$$x_i' = \frac{w^T x_i}{w^T w} w$$

- If w be a unit vector, that is, $w^T w = 1$,then

$$x_i' = (w^T x_i) w = \mathrm{a_i w}\,.$$

Projection

Iris dataset with sepal length and sepal width as the attributes, and iris-setosa as class c_1 (circles), and the other two Iris types as class c_2 (triangles). There are n_1 = 50 points in c_1 and n_2 = 100 points in c_2

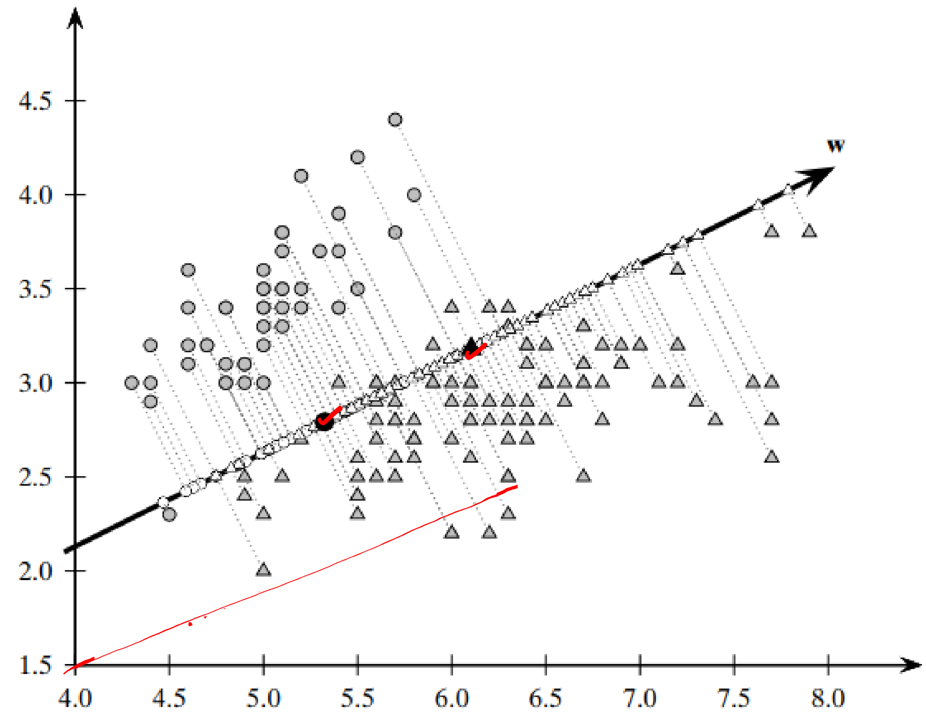


Figure 20.1. Projection onto **w**.

Mean of projected points

- For all data point in class c_1

$$m_1 = \frac{1}{n_1} \sum_{x_i \in C_1} w^T x_i = w^T \left(\frac{1}{n_1} \sum_{x_i \in C_1} x_i \right) = wT\mu_1$$

- Similarly for all data point in class c_2

$$m_2 = wT\mu_2$$

Mean of projected points

- For all data point in class c_1

$$m_1 = \frac{1}{n_1} \sum_{x_i \in C_1} w^T x_i = w^T \left(\frac{1}{n_1} \sum_{x_i \in C_1} x_i \right) = wT\mu_1$$

- Similarly for all data point in class c_2

$$m_2 = wT\mu_2$$

- The mean of the projected points is the same as the projection of the mean.

Mean of projected points

- For all data point in class c_1

$$m_1 = \frac{1}{n_1} \sum_{x_i \in C_1} w^T x_i = w^T \left(\frac{1}{n_1} \sum_{x_i \in C_1} x_i \right) = wT\mu_1$$

- Similarly for all data point in class c_2

$$m_2 = wT\mu_2$$

- The mean of the projected points is the same as the projection of the mean.

A good separation

- To maximize the separation between the classes, it seems reasonable to maximize the difference between the projected means, $|m_1 - m_2|$.

But is it enough ??

▪

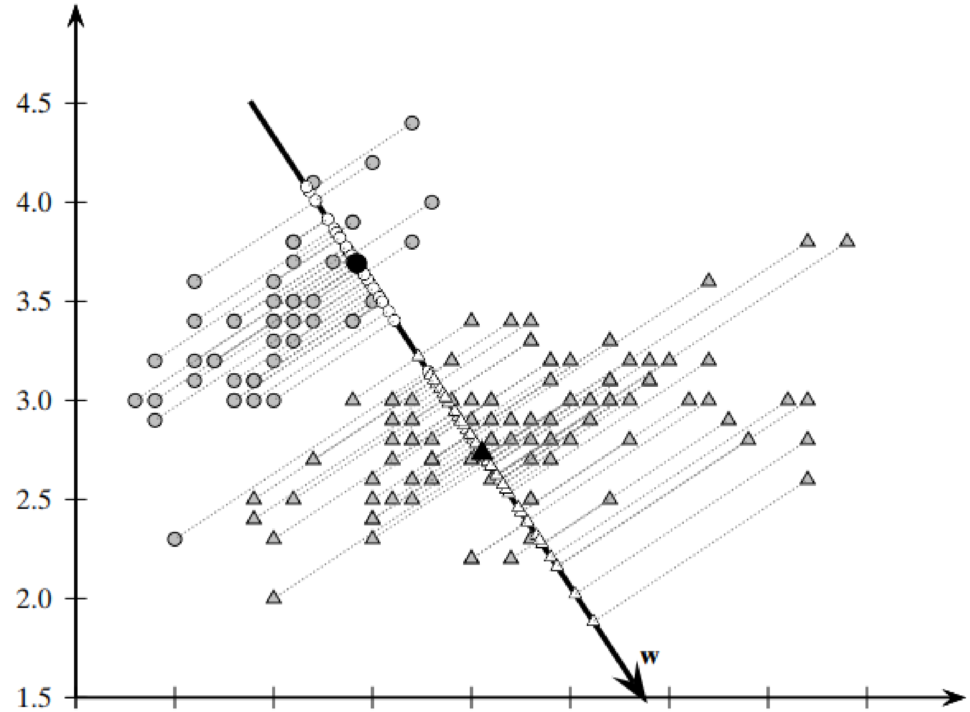


Figure 20.2. Linear discriminant direction **w**.

A good separation

- To maximize the separation between the classes, it seems reasonable to maximize the difference between the projected means, $|m_1 - m_2|$.

But is it enough ??

- For good separation, the variance of the projected points for each class should also not be too large. A large variance would lead to possible overlaps among the points of the two classes due to the large spread of the points, and thus we may fail to have a good separation.

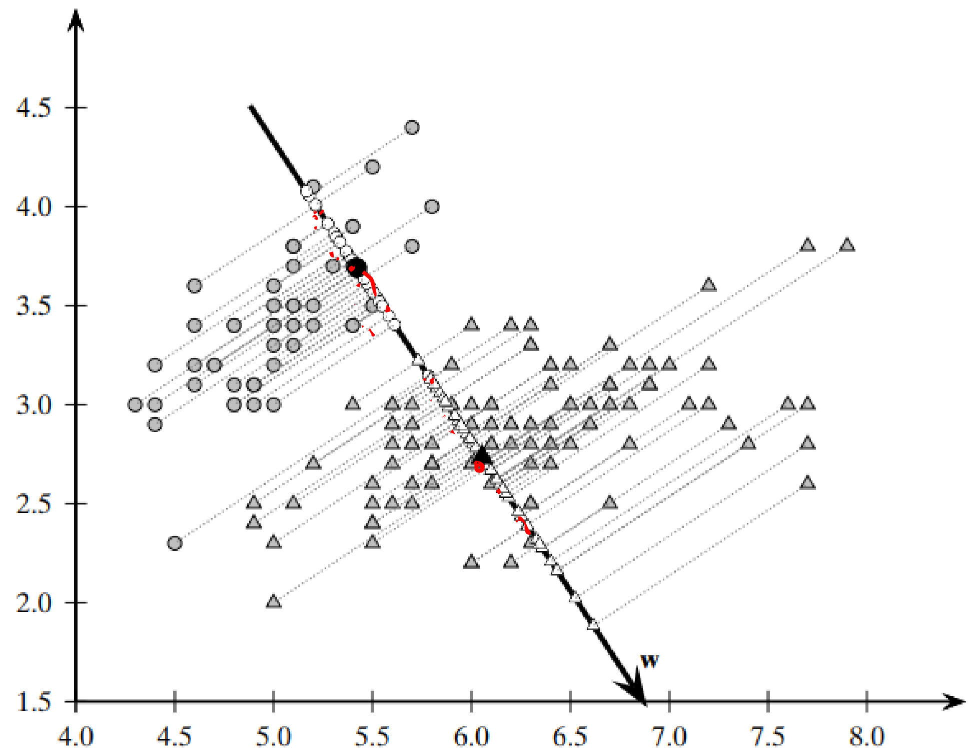


Figure 20.2. Linear discriminant direction w.

Scatter for projected data points

- LDA maximizes the separation by ensuring that the scatter s_j for the projected points within each class is small, where scatter is defined as

$$s_j = \sum_{x_i \in Cj} (ai - mj)^2$$

- Scatter is the total squared deviation from the mean, as opposed to the variance, which is the average deviation from mean.

Scatter for projected data points

- LDA maximizes the separation by ensuring that the scatter s_j for the projected points within each class is small, where scatter is defined as

$$s_j = \sum_{x_i \in Cj} (ai - mj)^{2}$$

- Scatter is the total squared deviation from the mean, as opposed to the variance, which is the average deviation from mean.

For good separation

- We can incorporate the two LDA criteria, namely,

I. maximizing the distance between projected means and

II. minimizing the sum of projected scatter,

into a single maximization criterion called the

Fisher LDA objective:

$$\max_{w} J(w) = \frac{(m_1 - m_2)^2}{{s_1}^2 + {s_2}^2}$$

For good separation

- Fisher LDA objective:

$$\max_{w} J(w) = \frac{(m_1 - m_2)^2}{{s_1}^2 + {s_2}^2}$$

The goal of LDA is to find the vector w that maximizes J(w)

- It is the direction that maximizes the separation between the two means m_1 and m_2, and minimizes the total scatter $s_1^2 + s_2^2$ of the two classes. The vector w is also called the optimal Linear Discriminant (LD).

LDA Objective

$(AB)^T = B^T A^T$ $m_1 - m_2 = w^T(\mu_1 - \mu_2)$

- Fisher LDA objective:

$$\max_w J(w) = \frac{(m_1 - m_2)^2}{{s_1}^2 + {s_2}^2}$$

$$(m_1 - m_2)^2 = (\underbrace{w^T(\mu_1 - \mu_2)}_{\|w\|})^2 = \boxed{wT(\mu_1 - \mu_2)}(\mu_1 - \mu_2)^T w$$

$$= \underbrace{w^T B\, w}_{\|w\|^2} =$$

Here B is the $d \times d$ between class scatter matrix.

$B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ $d \times d$

LDA Objective

- Fisher LDA objective:

$$\max_{w} J(w) = \frac{(m_1 - m_2)^2}{{s_1}^2 + {s_2}^2}$$

As for the projected scatter for class c_1 we compute

$$s_1 = \sum_{x_i \in C_1} (a_i - m_1)^2$$

$$= \sum_{x_i \in C_1} (w^T x_i - w^T \mu_1)^2$$

$$= \sum_{x_i \in C_1} (w^T (x_i - \mu_1))^2$$

LDA Objective

ϕ A

$$
\begin{aligned}
s_1 &= \sum_{x_i \in C_1} (a_i - m_1)^2 \\
&= \sum_{x_i \in C_1} (w^T x_i - w^T \mu_1)^2 \\
&= \sum_{x_i \in C_1} \underline{(w^T (x_i - \mu_1))}^2 \\
&= \sum_{x_i \in C_1} (w^T (x_i - \mu_1)(x_i - \mu_1)^T w) \\
&= \cancel{\sum_{x_i \in C_1}} \underline{(w^T S_1 w)}
\end{aligned}
$$

$$S_1 = \sum_{x_i \in C_1} (x - \mu_1)(x - \mu_1)^T$$

LDA Objective

$$s_1 = \sum_{x_i \in C_1} (a_i - m_1)^2$$

$$= \sum_{x_i \in C_1} (w^T x_i - w^T \mu_1)^2$$

$$= \sum_{x_i \in C_1} (w^T (x_i - \mu_1))^2$$

$$= \sum_{x_i \in C_1} (w^T (x_i - \mu_1)(x_i - \mu_1)^T w)$$

$$= \sum_{x_i \in C_1} (w^T S_1 w)$$

Similarly

$$s_2 = \sum_{x_i \in C_2} (w^T S_2 w)$$

LDA Objective

$$
\begin{aligned}
s_1 &= \sum_{x_i \in C_1} (a_i - m_1)^2 \\
&= \sum_{x_i \in C_1} (w^T x_i - w^T \mu_1)^2 \\
&= \sum_{x_i \in C_1} (w^T (x_i - \mu_1))^2 \\
&= \sum_{x_i \in C_1} (w^T (x_i - \mu_1)(x_i - \mu_1)^T w) \\
&= \cancel{\sum_{x_i \in C_1}} (w^T S_1 w)
\end{aligned}
$$

For good separation

$$\frac{w^T \beta w / \|w\|^2}{(w^T s_1 w + w^T s_2 w)/\|w\|^2}$$

- Fisher LDA objective:

$$\max_{w} J(w) = \frac{}{\cancel{m_1 = m_2}} s_1^2 + s_2^2$$

$$\frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

The goal of LDA is to find the vector w that maximizes J(w)

- It is the direction that maximizes the separation between the two means m_1 and m_2, and minimizes the total scatter $s_1^2 + s_2^2$ of the two classes. The vector w is also called the optimal linear discriminant (LD).

$$\max_{w} \frac{w^T B w}{w^T S_1 w + w^T S_2 w} = \frac{w^T B w}{w^T (S_1 + S_2) w}$$

$$= \frac{w^T B w}{w^T S w}$$

$S = S_1 + S_2$ which denotes the within class scatter matrix for pooled data

$\underline{d \times d}$

$$\max_{w} J(w) = \frac{w^T B w}{w^T S w}$$

$$\max_{w} J(w) = \frac{w^T B w}{w^T S w}$$

$$\nabla_w J(w) = ?$$

$$\frac{d\left(\frac{f(x)}{g(x)}\right)}{dx} = \frac{f'(x)g(x) - g'(x)f(x)}{(g(x))^2}$$

$$\Rightarrow \quad \frac{2Bw(w^T S w) - 2Sw(w^T B w)}{(w^T S w)^2} = 0$$

$$\Rightarrow \quad Bw\,(\underbrace{w^T S w}) = Sw\,(\underbrace{w^T B w})$$

$$\Rightarrow \quad Bw = \left\{\frac{w^T B w}{w^T S w}\right\} Sw \Rightarrow Bw = \underline{J(w)}\, Sw$$

$$Bw = \lambda S$$

$$S^{-1}Bw = \underbrace{J(w)}_{\lambda}\, w$$

$$\Rightarrow \underline{(S^{-1}B)}\, w = \lambda \, w$$

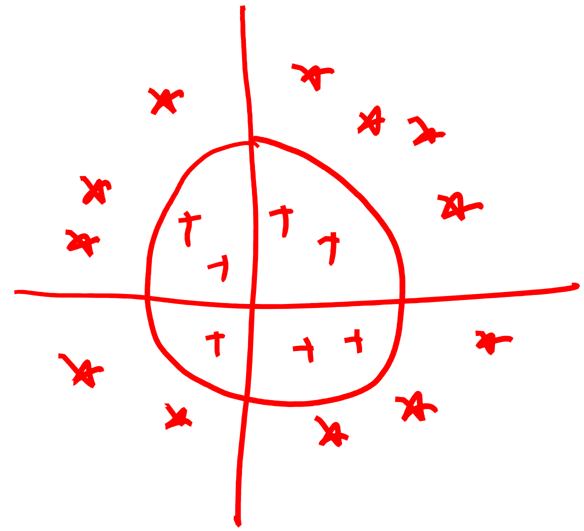
Eigenvector

Eigenvalue

To maximize $J(w)$ (λ), we need to look for largest eigenvalue of $(S^{-1}B)$

$\phi(x)$

x



$$u_1 = \frac{1}{n_1} \sum_{x_i \in C_1} \phi(x_i)$$

$$u_2 = \frac{1}{n_2} \sum_{x_i \in C_2} \phi(x_i)$$

$$= w^T (u_1 - u_2)(u_1 - u_2)^T w$$

$$(m_1 - m_2)^2$$

$$= \left[w^T (u_1 - u_2) \right]^2$$

$$(u_1 - u_2)(u_1 - u_2)^T$$

$$\left(\frac{1}{n_1}\sum_{i\in C_1}\varphi(x_i) - \frac{1}{n_2}\sum_{i\in C_2}\varphi(x_i)\right)\left(\frac{1}{n_1}\sum_{i\in C_1}\varphi(x_i) - \frac{1}{n_2}\sum_{i\in C_2}\varphi(x_i)\right)^T$$

$$\frac{1}{n_1^2}\left(\sum_{i\in C_1}\varphi(x_i)\right)\sum_{i\in C_1}\varphi(x_i)^T \rightarrow$$

$$- \frac{2}{n_1 n_2}\sum_{i\in C_1}\varphi(x_i)\sum_{i\in C_2}\varphi(x_i)^T \quad \frac{1}{n_2 n_1} + \frac{1}{n_2^2}\sum_{i\in C_2}\varphi(x_i)\sum_{i\in C_2}\varphi(x_i)$$

$$\frac{1}{n_1^2}\left(\phi(x_1)+\phi(x_2)+\cdots\phi(x_{n_1})\right)\left(\phi(x_1)+\phi(x_1)+\cdots\phi(x_{n_1})\right)^T$$

$$\phi(x_i)^T\phi(x_j) = \underline{K(x_i, x_j)}$$