Fisher LDA

Linear discriminant analysis (LDA) is a method to project data from 2 categories to a 1 dimensional space where they are "maximally separated."

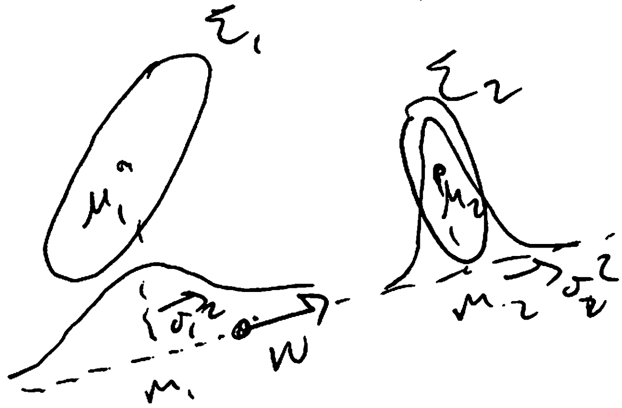
Fisher LDA measures seperability of the projections using $\frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)}$ where μ_1 and μ_2 are the mean values of the 2 projected data distributions; σ_1^2, σ_2^2 are the respective variances.

Suppose our data $x \in \mathbb{R}^n$ comes from 2 categories (pdfs) such that

$$x_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \quad \text{and} \quad x_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$$

Gaussian pdf   mean   covariance

Let $y_1 = w^T x_1$ and $y_2 = w^T x_2$ be the projected data when vector $w \in \mathbb{R}^n$ is used for this projection.



$E[y_1] = E[w^T x_1] = w^T E[x_1] = w^T \mu_1$

Similarly $E[y_2] = w^T \mu_2$

$Var(y_1) = \sigma_1^2 = E[(y_1 - \mu_1)^2] = E[(w^T x_1 - w^T \mu_1)^2]$

$= E[w^T (x_1 - \mu_1)(x_1 - \mu_1)^T w] = w^T E[(x_1 - \mu_1)(x_1 - \mu_1)^T] w$

$= w^T \Sigma_1 w$

Similarly $Var(y_2) = w^T \Sigma_2 w$

Fisher LDA looks for a w that maximizes the following objective function.

$$\arg\max_w \frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)} = \arg\max_w \frac{w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w}{w^T (\Sigma_1 + \Sigma_2) w}$$

Let $S_B \triangleq (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$ Between class scatter matrix

$S_W = (\Sigma_1 + \Sigma_2)$ within class scatter matrix

$$\therefore \max_{w} \frac{w^T S_B w}{w^T S_W w}$$

Take the gradient and equate to zero...

$$0 = \frac{\partial \left(\frac{w^T S_B w}{w^T S_W w} \right)}{\partial w} = \frac{2 S_B w (w^T S_W w) - 2 S_W w (w^T S_B w)}{(w^T S_B w)^2}$$

or equivalently $S_B w (w^T S_W w) = S_W w (w^T S_B w)$

or $S_B w = \left(\frac{w^T S_B w}{w^T S_W w} \right) S_W w$

$\therefore$ The optimal w is a generalized eigenvector of the matrix pair (S_B, S_W) that corresponds to the largest generalized eigenvalue $\left(\frac{w^T S_B w}{w^T S_W w} \right)$.

Alternatively, $S_W^{-1} S_B w = \left(\frac{w^T S_B w}{w^T S_W w} \right) w$.

$\therefore$ The optimal w is the eigenvector of $S_W^{-1} S_B$ that corresponds to the largest eigenvalue of this matrix.