LDA

Summary

LDA is the techniques for dimensionality reduction or liner classification. LDA assumes that the independent variables are normally distributed. LDA explicitly attemps to model the difference between the classes of data.

Algorithm

Within-class varianves should be small and between-class variance should be large. Consider transformed data $y = a^T x$. Maxmize the ratio of between-class scattter to within-class scatter (class ω_c):

$$J(a) = \frac{a^T S_B a}{a^T S_W a}, \tag{1}$$

where $S_B = (m_1 - m_2)(m_1 - m_2)^T$, $S_W = S_1 + S_2$, $S_c = \sum_{x \in \omega_c} (x - m_c)(x - m_c)^T$, $m = E(x)$. We want to know only the direction of a:

$$a \propto S_W^{-1} \left(m_1 - m_2 \right). \tag{2}$$

Examples

Consider two Gaussian distributions whose parameters are given as follows:

$$m_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \tag{3}$$

$$m_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \tag{4}$$

Apply Fisher LDA to the 2-d samples generated from th above two distributions by assuming the two distributions are two different classes. Draw the calculated axis on the scatter plot.