Linear Discriminant Analysis (线性判别分析)

(1) Model Assumption.

Let π_k be the prior probability of class k, i.e., $\sum_k \pi_k = 1$. Suppose $f_k(x)$ is the conditional density function of X given the class $G = k$ (we use G here to denote the class label). By Bayes Theorem,

$$P(G = k|X = x) = \frac{\pi_k f_k(x)}{\sum_k \pi_k f_k(x)}$$

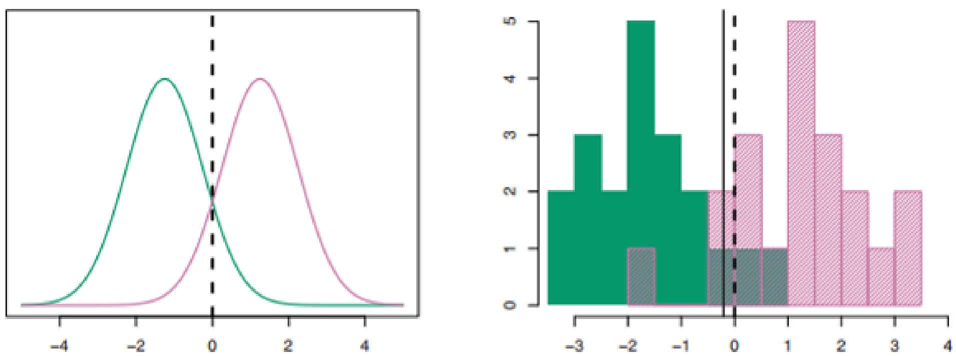

FIGURE 4.4. Left: *Two one-dimensional normal density functions are shown. The dashed vertical line represents the Bayes decision boundary.* Right: *20 observations were drawn from each of the two classes, and are shown as histograms. The Bayes decision boundary is again shown as a dashed vertical line. The solid vertical line represents the LDA decision boundary estimated from the training data.*

LDA uses Gaussian densities, i.e.,

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\Big\{ -\frac{1}{2}(x-\mu_k)^\top \Sigma_k^{-1}(x-\mu_k)\Big\} \tag{0.1}$$

The LDA assumes $\Sigma_k = \Sigma$. It suffices to look at the log-ratio,

$$\log \frac{P(G=k|X=x)}{P(G=l|X=x)} = \log \frac{f_k(x)}{f_l(x)} + \log \frac{\pi_k}{\pi_l},$$
$$= \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^\top \Sigma^{-1}(\mu_k - \mu_l) + (\mu_k - \mu_l)^\top \Sigma^{-1} x$$

(2) Linear discriminant functions:

$$\delta_k(x) = x^\top \Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^\top \Sigma^{-1}\mu_k + \log \pi_k.$$

(3) Parameter Estimation:

$$\widehat{\pi}_k = N_k/N, \quad \widehat{\mu}_k = \sum_{g_i=k} x_i/N_k, \quad \widehat{\Sigma} = \sum_{k=1}^{K}\sum_{g_i=k}(x_i - \widehat{\mu}_k)(x_i - \widehat{\mu}_k)^\top/(N-K)$$