Quadratic Discriminant Analysis Explained

James Taylor

We assume that our data come from k populations $\pi_1, \dots, \pi_k$ with respective prior probabilities $\gamma_1, \dots, \gamma_k$. We further assume that in the population π_i the pdf of $X = (X_1, \dots, X_p)$ is multivariate normal with mean vector μ_i and covariance matrix Σ_i. Thus,

$$f_i(x) := p(X = x \,|\, Y = i) = \frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right]$$

The QDA classifier assigns x to the population π_i for which $p(Y = i \,|\, X = x)$ is largest. By Bayes theorem,

$$\begin{aligned} p(Y = i \,|\, X = x) &= \frac{p(X = x \,|\, Y = i)p(Y = i)}{p(X = x)} \\ &= \frac{\gamma_i}{p(X = x)} f_i(x) \end{aligned}$$

Since $p(X = x)$ does not depend on i, it can be treated as a constant. So x will be assigned to the population π_i for which $\gamma_i f_i(x)$, or equivalently, $\log \gamma_i f_i(x)$ is largest. Now,

$$\begin{aligned} \log(\gamma_i f_i(x)) &= \log(f_i(x)) + \log(\gamma_i) \\ &= \log\left(\frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right]\right) + \log(\gamma_i) \\ &= \log\left(\frac{1}{(2\pi)^{p/2}|\Sigma_i|^{1/2}}\right) + \left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right] + \log(\gamma_i) \\ &= -\log\left((2\pi)^{p/2}|\Sigma_i|^{1/2}\right) + \left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right] + \log(\gamma_i) \\ &= -\log((2\pi)^{p/2}) - \log(|\Sigma_i|^{1/2}) + \left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right] + \log(\gamma_i) \end{aligned}$$

and since the first term on the last line doesn't depend on i, maximizing this expression is equivalent to maximizing

$$-\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + \log(\gamma_i)$$

Therefore, x will be assigned to the population π_i whose quadratic discriminant function

$$d_i^Q(x) = -\frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + \log(\gamma_i)$$

is largest when evaluated at x.

Note: We will need to estimate the prior probabilities γ_i, the population mean vectors μ_i, and the covariance matrices Σ_i. The latter two are estimated by the sample mean vectors and the sample covariance matrices, respectively.