Assignment Page 168.#3 105225007 吳家融

For case of LDA (linear discriminant analysis)

$X = (X_1, X_2, \dots, X_p)$ is drawn from a multivariate normal distribution.

$X \sim N(\mu, \Sigma)$. Here $\mathrm{E(X)} = \mu$ is the mean of X, and $\mathrm{Cov(X)} = \Sigma$ is the $\mathrm{p} \times \mathrm{p}$ covariance matrix of X. The multivariate normal density is defined as

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

$Y = 1,2,\dots,K$ class. $\pi_k = p(Y=k)$, $\Sigma_{k=1}^{K}\pi_k = 1$

$$f_k(\mathrm{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)\right)$$

Posterior probability of $\mathrm{Y = k}$

$$p_k(x) = P(\,Y = k \mid X = x\,) = \frac{\pi_k f_k(x)}{\Sigma_{l=1}^{k}\pi_l f_l(x)} \propto \pi_k f_k(x)$$

$$
\begin{aligned}
&log p_k(x) \\
&= log\pi_k + log f_k(x) \\
&= log\pi_k - \frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k) + const. \\
&= log\pi_k - \frac{1}{2}\{x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_k - \mu_k^T\Sigma^{-1}x + \mu_k^T\Sigma^{-1}\mu_k\} + const. \\
&= log\pi_k - \frac{1}{2}\{x^T\Sigma^{-1}x - 2\mu_k^T\Sigma^{-1}x + \mu_k^T\Sigma^{-1}\mu_k\} + const. \\
&= log\pi_k + \mu_k^T\Sigma^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k
\end{aligned}
$$

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

Data$(y_i, x_i), i = 1, \dots, n,$ where $x_i = (x_{i1}, \dots, x_{ip})^T$

The Bayes classifier for LDA = $\hat{\delta}_k\,(x) = log\hat{\pi}_k + \hat{\mu}_k^T\hat{\Sigma}^{-1}x - \frac{1}{2}\hat{\mu}_k^T\hat{\Sigma}^{-1}\hat{\mu}_k$

Where

$\hat{\pi}_k = \frac{n_k}{n}, n_k = \Sigma_{i,y_i=k}\, 1_{\{no\ of\ samples\ in\ class\ K\}}$

$\hat{\mu}_k = \frac{1}{n_k}\Sigma_{i,y_i=k}x_i$

$$\hat{\Sigma}_k = \frac{1}{n_k - 1}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$

$$= \frac{1}{n_k - 1}\Sigma_{i,y_{i=k}}\begin{pmatrix} (x_{i1-\hat{\mu}_{k1}})^2 & \cdots & (x_{i1} - \hat{\mu}_{k1})(x_{ip} - \hat{\mu}_{kp}) \\ \vdots & \ddots & \vdots \\ (x_{i1} - \hat{\mu}_{k1})(x_{ip} - \hat{\mu}_{kp}) & \cdots & (x_{i1-\hat{\mu}_{kp}})^2 \end{pmatrix}$$

$$\hat{\Sigma} = \frac{1}{n-k}\Sigma_{n=k}^{K}(n_k - 1)\hat{\Sigma}_k$$

For the case of QDA (quadratic discriminant analysis)

$$f_k(\mathrm{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}}\exp\left(-\frac{1}{2}(x - \mu_k)^T\Sigma^{-1}(x - \mu_k)\right)$$

$$\mathrm{P}_k(x) = P(Y = k \mid X = x) = \frac{\pi_k f_k(x)}{\Sigma_{l=1}^{k}\pi_l f_l(x)} \propto \pi_k f_k(x)$$

$$logp_k(x)$$

$$= \log(\pi_k) + logf_k(x)$$

$$= \ log\pi_k - \frac{1}{2}log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T\Sigma_k^{-1}(x - \mu_k) + const.$$

$$= \ log\pi_k - \frac{1}{2}log|\Sigma_k| - \frac{1}{2}\{x^T\Sigma_k^{-1}x - x^T\Sigma_k^{-1}\mu_k - \mu_k^T\Sigma_k^{-1}x + \mu_k^T\Sigma_k^{-1}\mu_k\} + const.$$

$$= \ log\pi_k - \frac{1}{2}log|\Sigma_k| - \frac{1}{2}\{x^T\Sigma_k^{-1}x - 2\mu_k^T\Sigma_k^{-1}x + \mu_k^T\Sigma_k^{-1}\mu_k\} + const.$$

$$= log\pi_k - \frac{1}{2}log|\Sigma_k| - x^T\Sigma_k^{-1}x + \mu_k^T\Sigma_k^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + const.$$

$$\delta_k(x) = \ log\pi_k - \frac{1}{2}log|\Sigma_k| - \boldsymbol{x^T\Sigma_k^{-1}x} + \mu_k^T\Sigma_k^{-1}x - \frac{1}{2}\mu_k^T\Sigma_k^{-1}\mu_k$$

The Bayes classifier for QDA

$$= \ \hat{\delta}_k\ (x) = \ log\hat{\pi}_k - \frac{1}{2}log\left|\hat{\Sigma}_k\right| - \boldsymbol{x^T\hat{\Sigma}_k^{-1}x} + \hat{\mu}_k^T\hat{\Sigma}_k^{-1}x - \frac{1}{2}\hat{\mu}_k^T\hat{\Sigma}_k^{-1}\hat{\mu}_k$$

Because of the term $\boldsymbol{x^T\hat{\Sigma}_k^{-1}x}$, the Bayes′ classifier is not linear and it is exact quadratic.