1. LDA and Logistic Regression

$$\log\left(\frac{p_1(x)}{1-p_1(x)}\right)=\log\left(\frac{p_1(x)}{p_2(x)}\right)$$

From 4.12:

$$p_k(x)=\frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K}\pi_l \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

$$\log\left(\frac{p_1(x)}{p_2(x)}\right)=\log\left[\frac{\pi_1 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_1)^2\right)}{\pi_2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_2)^2\right)}\right]$$

Rule 1
$\log(a*b)=\log(a)+\log(b)$

Rule 2
$\frac{c^d}{c^f}=c^{d-f}$

$$=\log\left(\frac{\pi_1}{\pi_2}\right)+\log\left(\exp\left(-\frac{1}{2\sigma^2}(x-\mu_1)^2+\frac{1}{2\sigma^2}(x-\mu_2)^2\right)\right)$$

$\log_e e^{\gamma}=\gamma$

$$=\log\left(\frac{\pi_1}{\pi_2}\right)+\frac{1}{2\sigma^2}\left[-(x-\mu_1)^2+(x-\mu_2)^2\right]$$

$$=\log\left(\frac{\pi_1}{\pi_2}\right)+\frac{1}{2\sigma^2}\left[\cancel{-x^2}+2\mu_1x-\mu_1^2+\cancel{x^2}-2\mu_2x+\mu_2^2\right]$$

cancelling out

$$=\log\left(\frac{\pi_1}{\pi_2}\right)+\frac{1}{2\sigma^2}\left[2x(\mu_1-\mu_2)+\mu_2^2-\mu_1^2\right]$$

$$=\underbrace{\log\left(\frac{\pi_1}{\pi_2}\right)+\frac{\mu_2^2-\mu_1^2}{2\sigma^2}}_{c_0}+\underbrace{(\mu_1-\mu_2)}_{c_1}x$$

Thus, $\log\left(\frac{p_1(x)}{p_2(x)}\right)=c_0+c_1x$ by direct proof ∎

2. Implementation of QDA

(a) Derive equation (4.23) on page 149 of ISL i.e.

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

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

Suppose that x is a vector $(x_1, x_2, \ldots, x_p)$ drawn from a multivariate normal distribution with mean μ_k and covariance matrix Σ_k for the k^{th} class

Then

$$f_x(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right)$$ Let this be

$P(X=x|Y=k)$. Assuming that the prior distribution defined by the equation $P(Y=k)=\pi_k$ is known,

$P(Y=k|X=k) = \frac{f_k(x)\pi_k}{P(X=x)}$ which equals $cf_k(x)\pi_k$. In this case, c is a constant. The equation expands to:

$$c\pi_k \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right)$$

Applying the logarithm:

$$\log(P_k(x)) = \log c - \log\left((2\pi)^{p/2}|\Sigma_k|^{\frac{1}{2}}\right) - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)$$

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

To maximize the above, we must ignore constant terms.

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

Thus, we have derived 4.23 ∎

2b). Assume that you have the data $\{(x_i, y_i)\}_{1\le i\le n}$. Write down your estimators of Σ_k, μ_k and π_k.

To estimate the Prior Probability of the k^{th} class, we use a sample size of it. ($\hat{\pi}_k = \frac{n_k}{n}$ where n is the total number of observations).

We apply the Principle of Maximum Likelihood Estimation the mean of this arbitrary class

$$\hat{\mu}_k = \frac{1}{\pi_k}\sum x_i\, I(c(x_i)=k).$$

Here $I = \begin{cases} 1 \text{ if element present} \\ 0 \text{ otherwise} \end{cases}$

In Quadratic Discriminant Analysis, the Covariance Matrix of our class is estimated as follows:

$$\Sigma_k^{-1} = \frac{1}{\pi_k}\sum_{i=1}^{n}(x_i - \hat{\mu}_k)^T(x_i - \mu_k)\, I(c(x_i)=k).$$

2c) Based on your estimators in (b), write down the decision rule of QDA.

The decision boundary between classes k and ℓ is $(x: r_k(x) : r_\ell(x))$ on

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

$$= \frac{1}{2}\log(|\Sigma_L|) - \frac{1}{2}x^T \Sigma_L^{-1} x - \frac{1}{2}\mu_L^T \Sigma_L^{-1} \mu_L + \mu_L^T \Sigma_L^{-1} x + \log(\pi_L)$$

QDA assigns observations for those classes that are largest as described in Part a.