Stat/CS 5525 Homework 3

Kazi Hasan Ibn Arif / hasanarif@vt.edu / SID: 906614469

12 October 2023

Answer to Ex. 4.2 (a)

To begin, the log-odds for classifying between the two classes using the LDA rule. Given Gaussian assumptions and a shared covariance matrix Σ, the log-odds of belonging to class 2 over class 1 for a given observation x is –

$$\log\left(\frac{Pr(G=2|X=x)}{Pr(G=1|X=x)}\right) = \log\left(\frac{\pi_2 \cdot \phi(x;\mu_2,\Sigma)}{\pi_1 \cdot \phi(x;\mu_1,\Sigma)}\right)$$

Where π_k represents the prior probability of class k and $\phi(x;\mu_k,\Sigma)$ represents the Gaussian density function for class k.

Expanding this equation, we get:

$$\log\left(\frac{\pi_2 \cdot \phi(x;\mu_2,\Sigma)}{\pi_1 \cdot \phi(x;\mu_1,\Sigma)}\right) = \log\left(\frac{\pi_2}{\pi_1}\right) - \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) + \frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

Simplifying, we get:

$$x^T\Sigma^{-1}(\mu_2-\mu_1) > \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\left(\frac{\pi_1}{\pi_2}\right)$$

Finally, plugging in the given class sizes $N1$ and $N2$ and the total number of samples N:

$$x^T\hat{\Sigma}^{-1}(\hat{\mu}_2-\hat{\mu}_1) > \frac{1}{2}\hat{\mu}_2^T\hat{\Sigma}^{-1}\hat{\mu}_2 - \frac{1}{2}\hat{\mu}_1^T\hat{\Sigma}^{-1}\hat{\mu}_1 + \log\left(\frac{N1}{N}\right) - \log\left(\frac{N2}{N}\right)$$

This proves that with the given target coding and class sizes, LDA classifies to class 2 if the inequality holds and to class 1 otherwise.

Answer to Ex. 4.2 (b)

We aim to minimize the least squares criterion:

$$J(\beta,\beta_0) = \sum_{i=1}^{N}\left(y_i - \beta_0 - \beta^T x_i\right)^2$$

The partial derivatives with respect to β and β_0 are set to zero:

1. For β:

$$\frac{\partial J}{\partial \beta} = -2X^T(\mathbf{y} - \beta_0 \mathbf{1}_N - X\beta) = 0$$

   Simplifying, we have:

$$\hat{\Sigma}\beta = (\hat{\mu}_2 - \hat{\mu}_1) - \beta_0 \mathbf{1}_p$$

2. For β_0:

$$\frac{\partial J}{\partial \beta_0} = -2\sum_{i=1}^{N}(y_i - \beta_0 - \beta^T x_i) = 0$$

   Simplifying, we get:

$$\beta_0 = \frac{1}{N}\left(\sum_{i=1}^{N} y_i - \sum_{i=1}^{N} \beta^T x_i\right)$$

After simplifying these equations, we get the final form:

$$(N-2)\hat{\Sigma}\beta + \frac{N_1 N_2}{N}\hat{\Sigma}_B\beta = N(\hat{\mu}_2 - \hat{\mu}_1)$$

where $\hat{\Sigma}_B = (\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$.

Answer to Ex. 4.2 (c)

To show that $\hat{\Sigma}_B\beta$ is in the direction $(\hat{\mu}_2 - \hat{\mu}_1)$ and $\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$.

Firstly, $\hat{\Sigma}_B\beta = (\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T\beta$ can be rewritten as:

$$\hat{\Sigma}_B\beta = (\hat{\mu}_2 - \hat{\mu}_1)((\hat{\mu}_2 - \hat{\mu}_1)^T\beta)$$

which shows $\hat{\Sigma}_B\beta$ is in the direction $(\hat{\mu}_2 - \hat{\mu}_1)$.

To show $\hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$, we use the formula for the inverse of a matrix of the form $(A + bb^T)$:

$$(A + bb^T)^{-1} = A^{-1} - \frac{A^{-1}bb^T A^{-1}}{1 + b^T A^{-1} b}$$

Applying this formula to the equation $(N-2)\hat{\Sigma} + \frac{N_1 N_2}{N}\hat{\Sigma}_B\beta = N(\hat{\mu}_2 - \hat{\mu}_1)$ it's straightforward to show that $\hat{\beta}$ is proportional to $\hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$.