STAT 524
HW8

Satoshi Ido
34788706
11/28/2023

11.1. Consider the two data sets

$$\mathbf{X}_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

for which

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

and

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) Calculate the linear discriminant function in (11-19).

(b) Classify the observation $\mathbf{x}_0' = [2 \quad 7]$ as population π_1 or population π_2, using Rule (11-18) with equal priors and equal costs.

(a) $\hat{y} = \hat{a}^T X$

$= [\bar{x}_1 - \bar{x}_2]^T S_{pooled} X$

$= \left[\begin{bmatrix} 3 \\ 6 \end{bmatrix} - \begin{bmatrix} 5 \\ 8 \end{bmatrix} \right]^T \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= [-2 \ \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad = -2x_1$

$\Leftrightarrow \hat{a}^T = [-2 \ \ 0] \qquad \therefore \ \underline{\hat{y} = -2X_1}$ //

(b) $\bar{y}_1 = [-2 \ \ 0] \begin{bmatrix} 3 \\ 6 \end{bmatrix} = -6$

$\bar{y}_2 = [-2 \ \ 0] \begin{bmatrix} 5 \\ 8 \end{bmatrix} = -10$ $\qquad \Rightarrow \hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = -8$

Then, we will allocate
$$\begin{cases} X_0 \text{ to } \pi_1 \text{ if } \hat{y}_0 \geq \hat{m} \\ X_0 \text{ to } \pi_2 \text{ if } \hat{y}_0 < \hat{m} \end{cases}$$

$\hat{y}_0 = [-2 \ \ 0] \begin{bmatrix} 2 \\ 7 \end{bmatrix} = -4 > -8$

$\therefore$ We will classify the observation X_0 as population π_1 //

11.5. Show that

$$-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)+\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}-\tfrac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

[see Equation (11-13).]

$$
\begin{aligned}
lhs &= \frac{-1}{2}\Sigma^{-1}\left[(X-\mu_1)^T(X-\mu_1)-(X-\mu_2)^T(X-\mu_2)\right]\\
&= \frac{-1}{2}\Sigma^{-1}\left[(X-\mu_1)^{2T}-(X-\mu_2)^{2T}\right]\\
&= \frac{-1}{2}\Sigma^{-1}\left[(X^2-2\mu_1X+\mu_1^2)^T-(X^2-2\mu_2X+\mu_2^2)^T\right]\\
&= \frac{-1}{2}\Sigma^{-1}\left[(\mu_1^2-\mu_2^2)-2X(\mu_1-\mu_2)^T\right]\\
&= \frac{-1}{2}\Sigma^{-1}\left[(\mu_1-\mu_2)^T(\mu_1+\mu_2)-2X(\mu_1-\mu_2)^T\right]\\
&= \frac{-1}{2}(\mu_1-\mu_2)\Sigma^{-1}(\mu_1+\mu_2)+(\mu_1-\mu_2)^T\Sigma^{-1}X\\
&= (\mu_1-\mu_2)\Sigma^{-1}X-\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1+\mu_2)\\
&= rhs
\end{aligned}
$$

Q.E.D

11.24. Annual financial data are collected for bankrupt firms approximately 2 years prior to their bankruptcy and for financially sound firms at about the same time. The data on four variables, X_1 = CF/TD = (cash flow)/(total debt), X_2 = NI/TA = (net income)/(total assets), X_3 = CA/CL = (current assets)/(current liabilities), and X_4 = CA/NS = (current assets)/(net sales), are given in Table 11.4.

(a) Using a different symbol for each group, plot the data for the pairs of observations (x_1, x_2), (x_1, x_3) and (x_1, x_4). Does it appear as if the data are approximately bivariate normal for any of these pairs of variables?

(b) Using the $n_1 = 21$ pairs of observations (x_1, x_2) for bankrupt firms and the $n_2 = 25$ pairs of observations (x_1, x_2) for nonbankrupt firms, calculate the sample mean vectors $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ and the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$.

(c) Using the results in (b) and assuming that both random samples are from bivariate normal populations, construct the classification rule (11-29) with $p_1 = p_2$ and $c(1|2) = c(2|1)$.

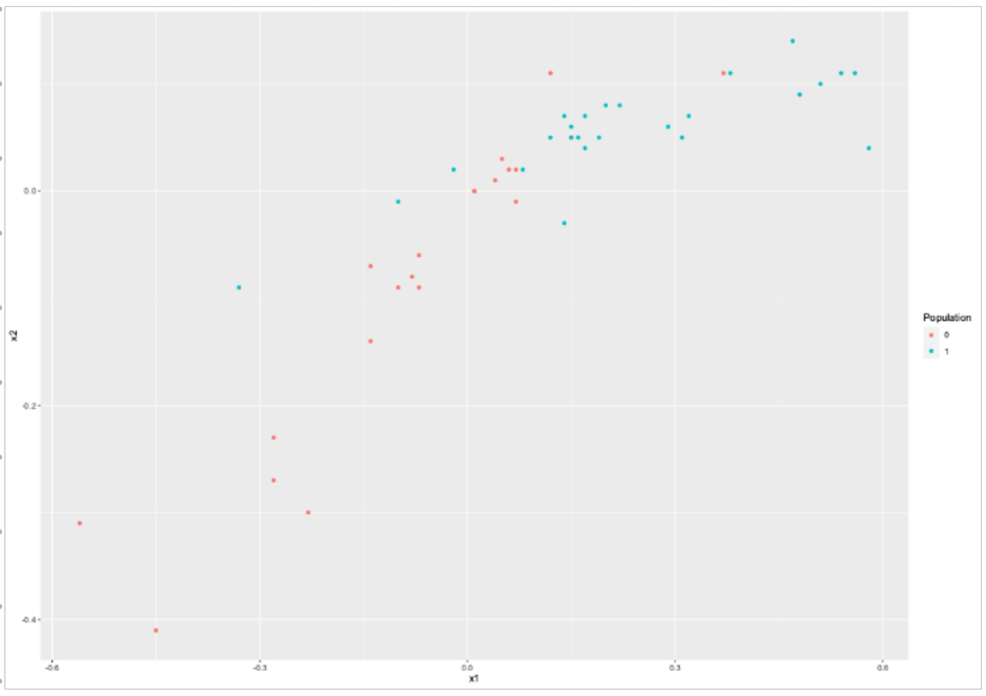
(d) Evaluate the performance of the classification rule developed in (c) by computing the apparent error rate (APER) from (11-34) and the estimated expected actual error rate $\hat{E}$ (AER) from (11-36).

(e) Repeat Parts c and d, assuming that $p_1 = .05$, $p_2 = .95$, and $c(1|2) = c(2|1)$. Is this choice of prior probabilities reasonable? Explain.

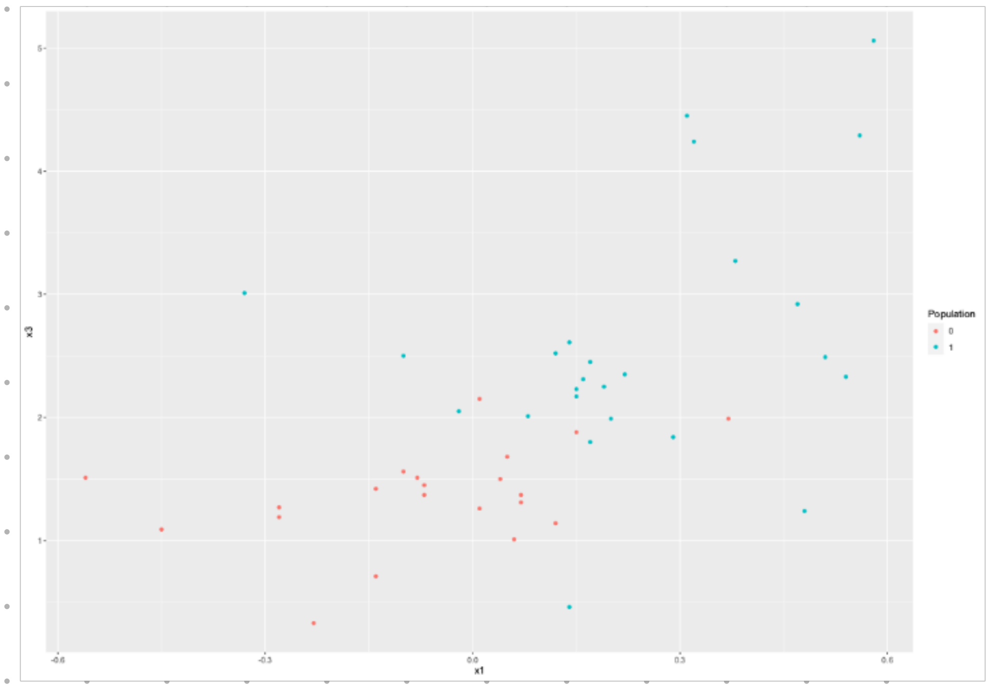
(f) Using the results in (b), form the pooled covariance matrix $\mathbf{S}_{\text{pooled}}$, and construct Fisher's sample linear discriminant function in (11-19). Use this function to classify the sample observations and evaluate the APER. Is Fisher's linear discriminant function a sensible choice for a classifier in this case? Explain.

I mainly used R to calculate those answers below.

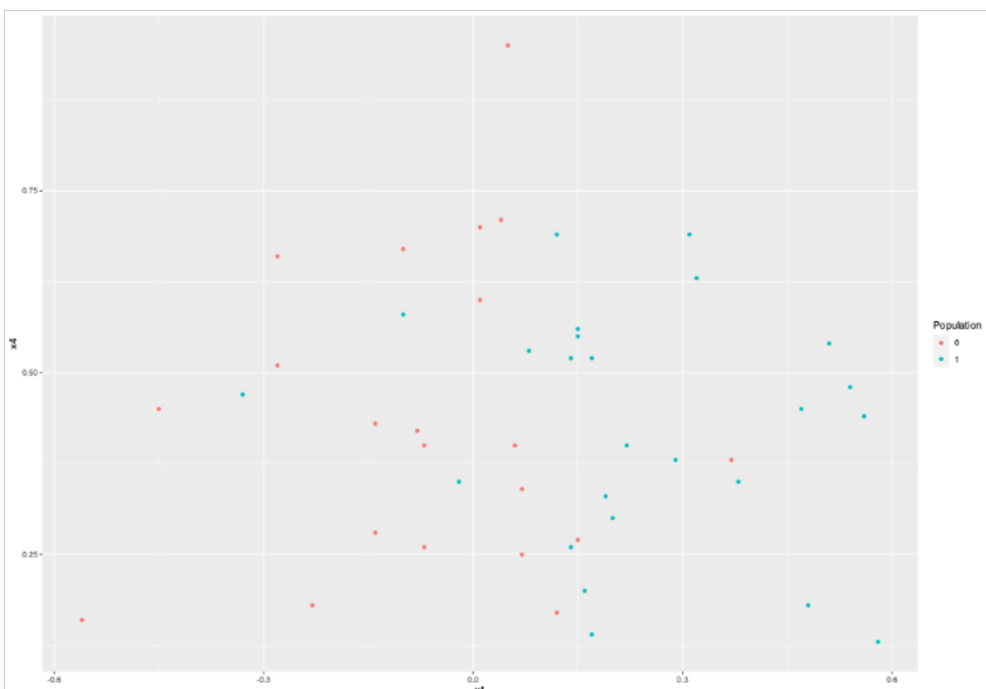
(a) Plot for (x_1, x_2)



Plot for (x_1, x_3)



Plot for (x_1, x_4)



From all these plots, the data points seem to lie in elliptical areas. Hence, the data approximately bivariate normal for these pairs. //

(b) $$\bar{x}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j} = \begin{bmatrix} -0.069 \\ -0.081 \end{bmatrix} \quad //$$

$$\bar{x}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j} = \begin{bmatrix} 0.235 \\ 0.056 \end{bmatrix} \quad //$$

$$S_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)^T$$

$$= \begin{bmatrix} 0.044 & 0.029 \\ 0.029 & 0.021 \end{bmatrix} \quad //$$

$$S_2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)^T$$

$$= \begin{bmatrix} 0.047 & 0.009 \\ 0.009 & 0.002 \end{bmatrix} //$$

(c) $P_1 = P_2 = 0.5$

$$\ln\left[\left(\frac{C(1|2)}{C(2|1)}\right)\left(\frac{P_2}{P_1}\right)\right] = 0$$

$$K = \frac{1}{2}\ln\left(\frac{|S_1|}{|S_2|}\right) + \frac{1}{2}(\mu_1^T \Sigma_1^{-1}\mu_1 - \mu_2^T \Sigma_2^{-1}\mu_2)$$

$$= 0.1624$$

$$\frac{-1}{2} X_0^T (S_1^{-1} - S_2^{-1}) X_0 + (\mu_1^T S_1^{-1} - \mu_2^T S_2^{-1}) X_0 - k \geq 0$$

$$= \frac{-1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 120.882 & -29.749 \\ -29.749 & -813.362 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 5.311 & -29.646 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$-0.162 \geq 0$$

$$= -60.441 X_1^2 + 29.749 X_1 X_2 + 406.681 X_2^2 + 5.311 X_1$$

$$-29.646 X_2 - 0.162 \geq 0$$ Quadratic function //

(d) By the quadratic function,

the confusion matrix :

		Predicted membership 0	1
Actual membership	0	13	8
	1	1	24

$$\therefore \text{APER} = \frac{8+1}{46} = 0.196 //$$

If we hold out the first observation

the confusion matrix :

		Predicted membership	
		0	1
Actual membership	0	13	8
	1	2	23

$\therefore \hat{E}(AER) = \frac{8+2}{46} = 0.232$ //

(f) $S_{pooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} = \begin{bmatrix} 0.046 & 0.018 \\ 0.018 & 0.011 \end{bmatrix}$ //

$\hat{y} = (\bar{X}_1 - \bar{X}_2)^T S^{-1}_{pooled} X = \hat{a}^T X$

$= -4.767 X_1 - 4.907 X_2$ //

$\hat{m} = -0.333$

We allocate X_0 to π_1 if $\hat{y}_0 \geq \hat{m}$

to π_2 if $\hat{y}_0 < \hat{m}$

the confusion matrix :

		Predicted membership	
		0	1
Actual membership	0	15	6
	1	3	22

$\therefore APER = \frac{6+3}{46} = 0.196$ //

$\therefore$ From the error rate, we can say that the linear discriminant is as good as the quadratic function is the case.

11.29. The GPA and GMAT data alluded to in Example 11.11 are listed in Table 11.6.

(a) Using these data, calculate $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \bar{\mathbf{x}}$, and $\mathbf{S}_{\text{pooled}}$ and thus verify the results for these quantities given in Example 11.11.

(b) Calculate $\mathbf{W}^{-1}$ and $\mathbf{B}$ and the eigenvalues and eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$. Use the linear discriminants derived from these eigenvectors to classify the new observation $\mathbf{x}_0' = [3.21 \quad 497]$ into one of the populations π_1: admit; π_2: not admit; and π_3: borderline. Does the classification agree with that in Example 11.11? Should it? Explain.

(a) From the R-output,

$$\bar{x}_1 = \begin{bmatrix} 3.404 \\ 561.226 \end{bmatrix} \qquad \bar{x}_2 = \begin{bmatrix} 2.483 \\ 447.071 \end{bmatrix} \qquad \bar{x}_3 = \begin{bmatrix} 2.993 \\ 446.231 \end{bmatrix}$$

$$\bar{x} = \begin{bmatrix} 2.975 \\ 488.447 \end{bmatrix} \qquad S_{pooled} = \begin{bmatrix} 0.036 & -2.019 \\ -2.019 & 3655.901 \end{bmatrix}$$

These are the same as eg 11.11 //

(b) $W = (n_1 + n_2 + n_3 - 3)\, S_{pooled}$

$$= \begin{bmatrix} 2.958 & -165.538 \\ -165.538 & 299783.892 \end{bmatrix}$$

$$W^{-1} = \begin{bmatrix} 0.349 & 0.000192 \\ 0.000192 & 0.0000344 \end{bmatrix}$$

$$B = \sum_{i=1}^{3} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T$$

$$= \begin{bmatrix} 0.427 & 50.839 \\ 50.839 & 8790.904 \end{bmatrix}$$

$$\therefore W^{-1}B = \begin{bmatrix} 0.159 & 19.431 \\ 0.000257 & 0.040 \end{bmatrix}$$

Eigen values : $\lambda_1 = 0.192$ Eigenvectors = $a_1^T = [0.999 \ 0.0017]$

$\lambda_2 = 0.007$ = $a_2^T = [-0.999 \ 0.0078]$

$$\hat{y}_1 = 0.9999 X_1 + 0.0017 X_2$$

$$\hat{y}_2 = -0.9999 X_1 + 0.0078 X_2$$

By using the formula,

$$\sum_{j=1}^{2} (\hat{y}_j - \bar{y}_{kj})^2 = \sum_{j=1}^{2} [a_j^T (X_0 - \bar{X}_i)]^2 , \ i = 1, 2, 3$$

Insert $X_0 = \begin{bmatrix} 3.21 \\ 497 \end{bmatrix}$ to this formula

$$\sum_{j=1}^{2} (\hat{y}_j - \bar{y}_{1j})^2 = (4.053 - 4.356)^2 + (0.667 - 0.975)^2$$

$$= 0.186$$

$$\sum_{j=1}^{2} (\hat{y}_j - \bar{y}_{2j})^2 = (4.053 - 3.241)^2 + (0.667 - 1.0056)^2$$

$$= 0.774$$

$$\sum_{j=1}^{2} (\hat{y}_j - \bar{y}_{3j})^2 = (4.053 - 3.750)^2 + (0.667 - 0.489)^2$$

$= 0.124$ This is the smallest.

$\therefore$ We allocate X_0 to π_3.

$\Rightarrow$ This is identical to eg 11.11. Any time, there are 3 populations with only 2 discriminants, using Fisher's Discriminants should be the same as using the sample distance method. //