s23 p-set 5 (For you use only. Please do not circulate or post.)

1. (a) i. 1-NN: $y = 1$ for $x \in [0, 0.6)$, $y = 0$ for $x \in (0.6, 1]$; $x = 0.6$ can be assigned to either class.
   ii. 3-NN: $y = 1$ for all $x \in [0, 1]$.

   (b) Average 2-NN: $y = 1$ for $x \in [0, 0.5]$, $y = 1/2$ for $x \in [0.5, 1]$.
   (here $x = 0.5$ as a training point with $y = 1$ is used; accept other decision with reasonable logic.)

2. (a) $\int_{-\infty}^{\infty} f_i(x)dx = 1$ results in $c_1 = 1, c_2 = 1, c_3 = 1/4$.

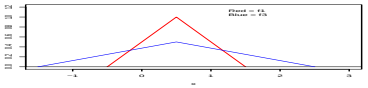
   (b) When $p_1 = 0.8, p_2 = 0.2$ and $c(1|2) = c(2|1)$, minimum ECM classification regions are

$$R_1 = \left\{x : \frac{f_1(x)}{f_2(x)} \geq \frac{p_2}{p_1} = \frac{1}{4}\right\}, \qquad R_2 = \left\{x : \frac{f_1(x)}{f_2(x)} < \frac{1}{4}\right\}$$

   The region divide is at $4f_1(x) = f_2(x)$, which yields $x = -1/3$. The resulted regions are

$$R_1 = \left\{x : -\frac{1}{3} < x \leq 1.5\right\}, \qquad R_2 = \left\{x : -1 \leq x \leq -\frac{1}{3}\right\}$$

   (c) Plot of f_1, f_3: When $p_1 = 0.8$, $p_3 = 0.2$ and $c(1|3) = c(3|1)$, minimum ECM classification regions are



$$R_1 = \{x : \frac{f_1(x)}{f_3(x)} \geq \frac{p_3}{p_1} = \frac{1}{4}\}, \qquad R_2 = \{x : \frac{f_1(x)}{f_3(x)} < \frac{1}{4}\}$$

   The region divide is at $4f_1(x) = f_3(x)$, which gives $x = 43/30, -13/30$. The resulted regions are

$$R_1 = \{x : -\frac{13}{30} \leq x \leq \frac{43}{30}\} \qquad R_2 = \{x : -1.5 \leq x < -\frac{13}{30}\} \cup \{x : \frac{43}{30} < x \leq 2.5\}$$

3. (a) The scatter plot ($\circ$ setosa, Δ versical, $+$ virginica) is shown in part (c).
   The points of all three groups appear to follow roughly ellipse-like patterns, thus bi-variate normal assumption is not unreasonable. However, the orientation of the ellipse appears to be different for the observations from π_1(Iris setosa), from the observation from π_2 and π_3. In π_1.
   There also appears to be an outlier (a setosa) closer to the group of π_2 (versicol) (labelled with a "$*$").

   (b) Assuming equal covariance matrices and normal populations, these are the linear discriminant scores $\hat{d}_i(\boldsymbol{x}) = \bar{\boldsymbol{x}}_i' \boldsymbol{s}_{pool}^{-1} \boldsymbol{x} - \frac{1}{2}\bar{\boldsymbol{x}}_i' \boldsymbol{s}_{pool}^{-1} \bar{\boldsymbol{x}}_i + \ln p_i$ for $i = 1, 2, 3$, with $\boldsymbol{x} = [\log Y_1, \log Y_2]'$.
   For both variables $logY_1$ and $logY_2$:

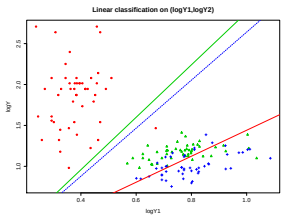
population	$\hat{d}_i(\log Y_1, \log Y_2)$
π_1 setosa	$26.81\, logY_1 + 28.90\, logY_2 - 31.97 + \ln(1/3)$
π_2 versical	$75.10\, logY_1 + 13.82\, logY_2 - 36.83 + \ln(1/3)$
π_1 virginica	$79.94\, logY_1 + 10.80\, logY_2 - 37.30 + \ln(1/3)$

   (c) Write $d_i = \hat{d}_i(\log Y_1, \log Y_2)$ in (b), the classification regions are formed by lines $d_i = d_j$. The top line is $d_2 = d_1$, the bottom line is $d_3 = d_2$. These two lines form the three classification regions in the plotted area.
   The border line (dashed) $d_1 = d_3$ in the plotted area is redundant: the classification of any point by the line would be re-evaluated by the other two partition lines. The line won't be redundant in expanded areas in $\mathbb{R}^2$.

   (d) The training error $APER$ and the holdout estimate of $\hat{E}(AER)$ (by cross-validation error) are in the table below.

Variables	train err $APER$	holdout-1 $\hat{E}(AER)$
$logY_1, logY_2$	$\frac{26}{150} = .17$	$\frac{27}{150} = .18$

   The preceding misclassification rates are not vey good. Using shape variables Y_1, Y_2 is effective in discriminating π_1 from π_2 and π_3, but not good at discriminating π_2 from π_3, because the overlap of π_1 and π_2 in both shape variables. Therefore, shape is not an effective discriminator of all three species of iris.



4. (a) $\bar{\boldsymbol{x}}_1 = [3.40, 561.32]', \bar{\boldsymbol{x}}_2 = [2.48, 447.07]', \bar{\boldsymbol{x}}_3 = [2.99, 446.23]'$, $\bar{\boldsymbol{x}} = [2.97, 488.45]$, $\boldsymbol{S}_{pool} = \begin{bmatrix} 0.036 & -2.019 \\ -2.019 & 3655.901 \end{bmatrix}$.

   (b) $\boldsymbol{W} = \begin{bmatrix} 2.958 & -165.538 \\ -165.538 & 299783.892 \end{bmatrix}$, $\boldsymbol{W}^{-1} = \begin{bmatrix} 0.348899 & 0.000193 \\ 0.000193 & .000003 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 0.426 & 50.678 \\ 50.678 & 8751.929 \end{bmatrix}$
   The eigenvalues and a pair of scaled eigenvectors of $\boldsymbol{W^{-1}B}$ are

$$\hat{\lambda}_1 = 0.191, \hat{\boldsymbol{a}}_1' = \begin{bmatrix} -5.009 \\ -0.009 \end{bmatrix}, \quad \hat{\boldsymbol{a}}_1' \boldsymbol{S}_{pool} \hat{\boldsymbol{a}}_1 \approx 1; \quad \hat{\lambda}_2 = 0.007, \hat{\boldsymbol{a}}_2' = \begin{bmatrix} 1.877 \\ -0.014 \end{bmatrix}, \quad \hat{\boldsymbol{a}}_2' \boldsymbol{S}_{pool} \hat{\boldsymbol{a}}_2 \approx 1 \quad (check\ \hat{\boldsymbol{a}}_1' \boldsymbol{S}_{pool} \hat{\boldsymbol{a}}_2 \approx 0)$$

   Notes: Here negative signs are used to be able to use the plot output of R function `lda` directly. You may have chosen $-\hat{\boldsymbol{a}}_1, -\hat{\boldsymbol{a}}_2$ as the eigenvectors, and the numerical values may differ slightly.

   (c) Allocation of new points. As instructed in the note, a common measure using discriminants is to allocate $\boldsymbol{x}$ to π_k if $\hat{\boldsymbol{a}}'\boldsymbol{x}$ is the closest to $\hat{\boldsymbol{a}}'\bar{\boldsymbol{x}}_k$ ($\hat{\boldsymbol{a}} = [\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2]$), that is, if

$$\|\hat{\boldsymbol{a}}'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)\|^2 = \sum_{j=1}^{2} [\hat{\boldsymbol{a}}_j'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)]^2 = \hat{\boldsymbol{a}}_1'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)^2 + \hat{\boldsymbol{a}}_2'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)^2 \leq \hat{\boldsymbol{a}}_1'(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)^2 + \hat{\boldsymbol{a}}_2'(\boldsymbol{x} - \bar{\boldsymbol{x}}_i)^2 \quad \text{for all } i = 1, 2, 3.$$

   It is equivalent to using the centered value $\boldsymbol{x} - \bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}$ in the place of $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}_i$ respectively.
   For $\boldsymbol{x} = [3.21, 497]'$,

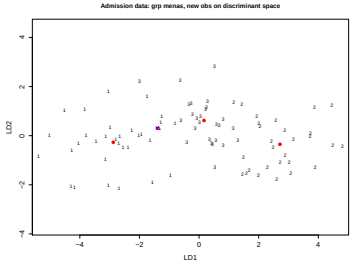
$$\sum_{j=1}^{2} [\hat{a}_j'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)]^2 = \begin{cases} 2.63, & k = 1 \\ 16.99, & k = 2 \\ 2.43, & k = 3 \end{cases}$$

   Thus $\boldsymbol{x} = [3.21, 497]'$ is assigned to π_3, the Borderline group. For $\boldsymbol{x} = [3.22, 497]'$,

$$\sum_{j=1}^{2} [\hat{a}_j'(\boldsymbol{x} - \bar{\boldsymbol{x}}_k)]^2 = \begin{cases} 2.50, & k = 1 \\ 17.43, & k = 2 \\ 2.57, & k = 3 \end{cases}$$

   Thus $\boldsymbol{x} = [3.22, 497]'$ is assigned to π_1, the Admit group.
   (Notice that this is a special case of three populations with only two discriminants, thus the classification results using Euclidean norm of Fisher's Discriminants will be equivalent to using Euclidean sample distance. However here the discriminant analysis practically reduced the dimension.)

   (d) The plots show (centered) observations in the coordinate system of discriminants 1 and 2, labeled by admissions (1=admit, 2=reject, 3=borderline). The group means are the solid disks. The new observation $[3.21, 397]$ is denoted by the triangle, the new observation $[3.22, 397]$ by `x`. The two applicants are similar with slight difference in GPA. The admission results are different. Not a good policy. GPA shouldn't be taken so seriously...



5. (Ref.: related plots are in Q6)

(a) The component means $\bar{\boldsymbol{x}}_1 = [3\ 6]'$, $\bar{\boldsymbol{x}}_2 = [5\ 8]'$. The linear discriminant function is

$$\hat{y} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}_{pool}^{-1} \boldsymbol{x} = \begin{bmatrix} -2 & 0 \end{bmatrix} \boldsymbol{x}$$

(b) $\bar{y}_1 = \hat{\boldsymbol{a}}' \bar{\boldsymbol{x}}_1 = -6$, $\bar{y}_2 = \hat{\boldsymbol{a}}' \bar{\boldsymbol{x}}_2 = -10$. $\hat{m} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}_{pool}^{-1} \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = -8$.

$$\text{Assign } \boldsymbol{x} \text{ to } \begin{cases} \pi_1, & \text{if } \hat{y} = \begin{bmatrix} -2 & 0 \end{bmatrix} \boldsymbol{x} \geq \hat{m}, \\ \pi_2, & \text{if } \hat{y} = \begin{bmatrix} -2 & 0 \end{bmatrix} \boldsymbol{x} < \hat{m}, \end{cases} \qquad \hat{m} = -8.$$

Since $\begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} 4.1 \\ 5 \end{bmatrix} = -8.2 < -8$, we assign $\boldsymbol{x} = [4.1\ 5]'$ to population π_2.

Since $\begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} 3.9 \\ 9 \end{bmatrix} = -7.8 > -8$, we assign $\boldsymbol{x} = [3.9\ 9]'$ to population π_1.

(c) Assign $\boldsymbol{x_0}$ to π_1 if

$$(\bar{\boldsymbol{x}}_{\boldsymbol{1}} - \bar{\boldsymbol{x}}_{\boldsymbol{2}})' \boldsymbol{S}_{\boldsymbol{pooled}}^{\boldsymbol{-1}} \boldsymbol{x_0} - \hat{m} \geq \log\left[\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}\right] = \log\left(\frac{20}{3} \times \frac{0.9}{0.1}\right) = 4.09$$

otherwise assign $\boldsymbol{x_0}$ to π_2.

$\begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} 4.1 \\ 5 \end{bmatrix} - \hat{m} = -8.2 - (-8) = -0.2 < 4.09$, we assign $\boldsymbol{x} = [4.1\ 5]'$ to population π_2.

$\begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} 3.9 \\ 9 \end{bmatrix} - \hat{m} = -7.8 - (-8) = 0.2 < 4.09$, we assign $\boldsymbol{x} = [3.9\ 9]'$ to population π_2 as well.

(d) Assign $\boldsymbol{x}$ to π_1 if

$$d(\boldsymbol{x}) = -\frac{1}{2} \boldsymbol{x}' (\boldsymbol{S}_{\boldsymbol{1}}^{-1} - \boldsymbol{S}_{\boldsymbol{2}}^{-1}) \boldsymbol{x} + (\bar{\boldsymbol{x}}_{\boldsymbol{1}}' \boldsymbol{S}_{\boldsymbol{1}}^{\boldsymbol{-1}} - \bar{\boldsymbol{x}}_{\boldsymbol{2}}' \boldsymbol{S}_{\boldsymbol{2}}^{\boldsymbol{-1}}) \boldsymbol{x} - \hat{k} \geq 0$$

otherwise assign $\boldsymbol{x_0}$ to π_2.

$d\left(\begin{bmatrix} 4.1 \\ 5 \end{bmatrix}\right) = -0.55 < 0$, we assign $\boldsymbol{x} = [4.1\ 5]'$ to population π_2

$d\left(\begin{bmatrix} 3.9 \\ 9 \end{bmatrix}\right) = -0.01 < 0$, we assign $\boldsymbol{x} = [3.9\ 9]'$ to population π_2 as well.

(e) In part (b), we assume equal variance matrix and assign $\boldsymbol{x} = [4.1\ 5]'$ to population π_2, but assign $\boldsymbol{x} = [3.9\ 9]'$ to population π_1.

In part (d), we assume $\Sigma_1 \neq \Sigma_2$ and normal assumption and assign both $\boldsymbol{x} = [4.1\ 5]'$ and $\boldsymbol{x} = [3.9\ 9]'$ to population π_2.

Because of the difference of the covariance assumptions for (b) and (d), we get different classification results (even when the costs and priors are the same). Therefore, the underlying assumptions for classification can be important and sometimes can markedly affect the results.

(f) The Hotelling's T^2 statistic is

$$T^2 = (\bar{\boldsymbol{x}}_{\boldsymbol{1}} - \bar{\boldsymbol{x}}_{\boldsymbol{2}})' \left[(\frac{1}{n_1} + \frac{1}{n_2}) \boldsymbol{S}_{\boldsymbol{pooled}}\right]^{-1} (\bar{\boldsymbol{x}}_{\boldsymbol{1}} - \bar{\boldsymbol{x}}_{\boldsymbol{2}})$$

Under H_0,

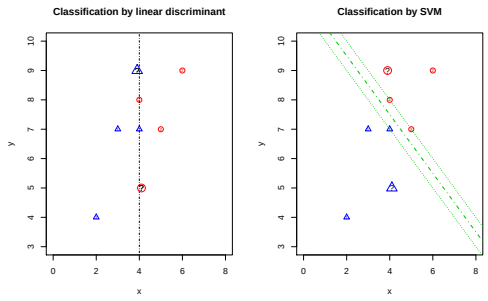
$$T^2 \sim \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1} = \frac{8}{3} F_{2,3}, \qquad \frac{8}{3} F_{2,3,.05} = 25.47$$

From the data, $T^2 = 6 < 25.47$. The data show no evidence against the null hypothesis that the population mean vectors are the same.

6. In the base plot, "Δ" for class 1, "$\bigcirc$" for class 2, "?" for points $(4.1, 5)$ and $(3.9, 9)$ to be classified.

(a) Assuming $\Sigma_1 = \Sigma_2$ the linear discriminant $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}_{pool}^{-1} \boldsymbol{x} = -8$, $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ is the line $x = 4$ in the x-y plane.

The new observations $(4.1, 5)$ and $(3.9, 9)$ are assigned to class 2 and 1, respectively, shown in the left plot.



(b) A (conceptual) SVM classifier and margins are shown on the right plot.

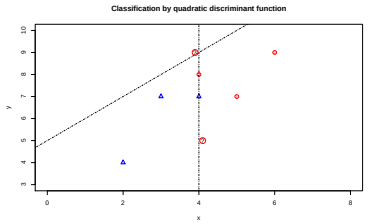
i. The SVM classifier and margins are shown on the right plot above (you may have a lightly different one).

ii. Plots (4,7) in class 1, (5,7) in class 2 (and (4,8) in class 2 by this classifier) are the supporting vectors.

iii. By SVM, $(4.1, 5)$ is assigned to class 1, $(3.9, 9)$ to class 2.

(c) Comparison of (b) and (c).
The classifiers are quite different: The two new points are assigned to opposite classes by the two methods. Both classifiers are valid. In this case, SVM seems more consistent with given data; more reasonable.

(d) The quadratic classifier is
$d(\boldsymbol{x}) = -\frac{1}{2} \boldsymbol{x}' (\boldsymbol{S}_1^{-1} - \boldsymbol{S}_2^{-1}) \boldsymbol{x} + (\bar{\boldsymbol{x}}_1' \boldsymbol{S}_1^{-1} - \bar{\boldsymbol{x}}_2' \boldsymbol{S}_2^{-1}) \boldsymbol{x} - \hat{k}$, $\hat{k} = \frac{1}{2} (\bar{\boldsymbol{x}}_1' \boldsymbol{S}_1^{-1} \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2' \boldsymbol{S}_2^{-1} \bar{\boldsymbol{x}}_2) + \frac{1}{2} \ln \frac{|\boldsymbol{S}_1|}{|\boldsymbol{S}_2|}$.



i. With $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$,

$$3d(\boldsymbol{x}) = 3d(x, y) = -4x(x - y) - 4x - 16y - 3k, \qquad k = -80/3.$$

The quadratic classifier's border $3d(\boldsymbol{x}) = 0$ (a degenerate hyperbolic conic section) can be factored as $4(4 - x)(x + 5 - y) = 0$, with $4(4 - x)(x + 5 - y) \geq 0$ for class 1, < 0 for class 2.
The border consists of two lines: $x = 4, y = x + 5$, as shown in the plot.

ii. Both observations $(4.1, 5)$ and $(3.9, 9)$ are assigned to class 2.

iii. The observation $(4.1, 9.5)$ will be classified into class 1. Not reasonable based on observed data.

iv. This time the quadratic rule is not better than the linear discriminant classifier in part (a), much less than the SVM classifier. The quadratic rule gives cross-over or overlapping classification regions not supported by the observed data.