תרגיל כיתה 5 – סיווג

שאלה 1 – כלל הסיווג

התייחס למקרה של 3 קבוצות (3 צורות) שמיוצגות על ידי שני מאפיינים. מטריצת הקווריאנס בשלוש הקבוצות זהה ואלכסונית. מרכזי הקבוצות שונים.

$$\Sigma_i = \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

$$\underline{\mu}_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \underline{\mu}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \quad \underline{\mu}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$P(w_1) = P(w_2) = P(w_3)$$

$$p(\underline{x}|w_i) \sim N(\underline{\mu}_i, \Sigma_i)$$

Shape of PDF Contours. Figure 6(a) shows the contours of the class-conditioned pdfs. Note that they are ellipses, as expected . For example, since

$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

computing $(\underline{x} - \underline{\mu}_i)^T \Sigma^{-1} (\underline{x} - \underline{\mu}_i) = k$, where k is an arbitrary (positive) constant, for each class yields the following:

$$\text{for } w_1 \;:\; x_1^2 + 2(x_2 - 2)^2 = k \text{ or } \left(\frac{x_1}{\sqrt{2}}\right)^2 + \left(\frac{x_2 - 2}{1}\right)^2 = \frac{k}{2}$$

$$\text{for } w_2 \;:\; (x_1 - 4)^2 + 2(x_2 - 1)^2 = k \text{ or } \left(\frac{x_1 - 4}{\sqrt{2}}\right)^2 + \left(\frac{x_2 - 1}{1}\right)^2 = \frac{k}{2}$$

$$\text{for } w_3 \;:\; (x_1 - 1)^2 + 2x_2^2 = k \text{ or } \left(\frac{x_1 - 1}{\sqrt{2}}\right)^2 + \left(\frac{x_2}{1}\right)^2 = \frac{k}{2}$$

Discriminant Functions.

$$g_1(\underline{x}) = (0 \quad 4)^T \underline{x} - 4 = 4x_2 - 4$$

$$g_2(\underline{x}) = (4 \quad 2)^T \underline{x} - 9 = 4x_1 + 2x_2 - 9$$

$$g_3(\underline{x}) = (1 \quad 0)^T \underline{x} - \frac{1}{2} = x_1 - \frac{1}{2}$$

Separating Hyperplanes. Since $c = 3$, there are $(c^2 - c)/2 = 3$ separating hyperplanes They are shown in Figure 6(b) and are computed as follows:

$$\underline{w}_{12} = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$$

with accompanying hyperplane $H_{12} : 4x_1 - 2x_2 - 5 = 0$

$$\underline{w}_{23} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

with accompanying hyperplane $H_{23} : 6x_1 + 4x_2 - 17 = 0$ and

$$\underline{w}_{31} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$$

with accompanying hyperplane $H_{31} : 2x_1 - 8x_2 + 7 = 0$. Superimposing Figure 6(b) onto Figure 6(a) yields an excellent visual display of the manner in which the separating hyperplanes account for the orientation of the pdfs. This is shown in Figure 6(c). ■

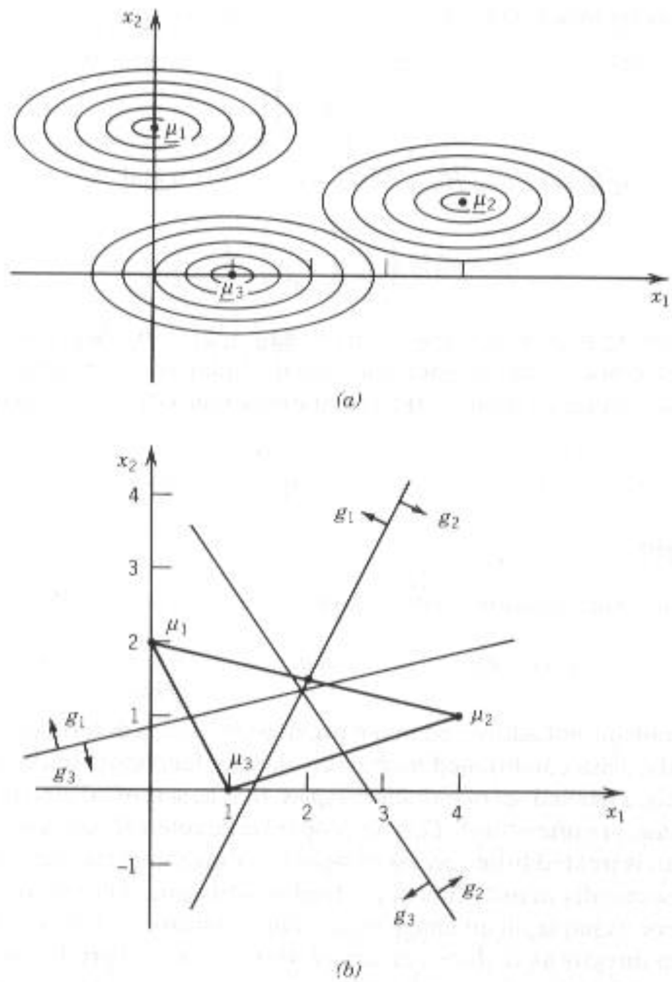

Figure 6: Example 5: Densities and decision regions.
(a) Visualization of class-conditioned pdf contours.
(b) Separating hyperplanes.

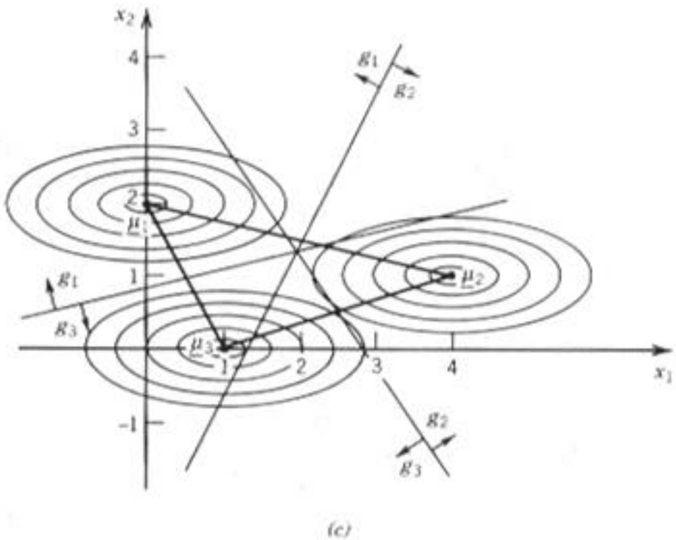

Figure 6 (cont.): **(c)** Superposition of decision boundaries of (b) on pdf contours of (a) to show partitioning of R^2.

<u>שאלה 2</u>

העתקה לא לינארית של מאפיינים :

התייחס למטלת זיהוי של שתי צורות שמיוצגות על ידי n מאפיינים. ההנחה היא שוקטרו המאפיינים מפולג גאוסית בשתי הקבוצות. כמו כן, וקטור המאפיינים הממוצע בשתי הקבוצות שווה לאפס.

מטריצת הקוואריאנס בשתי הקבוצות הן $\Sigma_1 = \sigma_1^2 I$ ו- $\Sigma_2 = \sigma_2^2 I$
הסיווג לפי הסבירות המירבית הוא במקרה זה לפי :

$$\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right)\sum_{i=1}^{n} x_i^2 + n\ln\left(\frac{\sigma_1^2}{\sigma_2^2}\right) \underset{C_2}{\overset{C_1}{\lessgtr}} 2\ln\left(\frac{p(C_1)}{p(C_2)}\right)$$

הסיווג תלוי במקרה זה במאפיין יחיד ששוה לסכום ריבוי המאפיינים המקוריים

$$y = \sum_{i=1}^{n} x_i^2 = \left|\underline{x}\right|^2$$

השימוש במאפיין יחיד זה שקול לשימוש בוקטור n המאפיינים המקורי (אותו כלל סיווג), ולכן גם שגיאת הסיווג זהה.

שאלה 3

העתקה של וקטור מאפיינים מפולג גאוסית:

העתקה לא סינגולרית A מעתיקה את וקטור המאפיינים המקורי $\underline{x}$, n-ממדי ומפולג גאוסית, לוקטור מאפיינים אחר n-ממדי $\underline{y}$.

כידוע, העתקה לינארית של וקטור מפולג גאוסית יוצרת וקטור שגם הוא מפולג גאוסית.

אם וקטור המאפיינים המקורי $\underline{x}$ הוא בעל תוחלת $\underline{\mu}_x$ ומטריצת קווריאנס Σ_x, אז הוקטור החדש $\underline{y}$ הוא בעל תוחלת $\underline{\mu}_y = A\underline{\mu}_x$ ומטריצת קווריאנס Σ_y שמקיימת:

$$\Sigma_x^{-1} = A^T \Sigma_y^{-1} A$$

הסיווג של וקטור המאפיינים המקורי $\underline{x}$, כאשר מטריצת הקווריאנס זהה בכל הקבוצות, הוא על ידי מזעור של

$$d_i(\underline{x}) = (\underline{x} - \underline{\mu}_i)^T \Sigma_x^{-1} (\underline{x} - \underline{\mu}_i)$$

הסיווג של וקטור המאפיינים החדש $\underline{y}$ הוא באופן דומה על ידי מזעור של

$$d_i(\underline{y}) = (\underline{y} - A\underline{\mu}_i)^T \Sigma_y^{-1} (\underline{y} - A\underline{\mu}_i) = (\underline{x} - \underline{\mu}_i)^T A^T \Sigma_y^{-1} A (\underline{x} - \underline{\mu}_i) = d_i(\underline{x})$$

כלל הסיווג זהה עבור שני וקטורי המאפיינים, ולכן גם שגיאת הסיווג זהה.

<u>שאלה 4</u> – שערוך הפרמטרים של פונקצית הפילוג

(א) יש לשערך את וקטור התוחלות של פילוג גאוסי על סמך דוגמאות, בהנחה שמטריצת הקווריאנס נתונה :

Here we assume $\underline{\mu}_i = \underline{\theta}_i = \underline{\mu}$, where

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

The training set sample $\underline{x}_k$ of H_i yields

$$\log\{p(\underline{x}_k|\underline{\mu})\} = -\frac{1}{2}(\underline{x}_k-\underline{\mu})^T\Sigma^{-1}(\underline{x}_k-\underline{\mu}) - \frac{1}{2}\log\{(2\pi)^d|\Sigma|\} \qquad (3-6)$$

Forming (see Appendix 1) the gradient of (3-6)

$$\nabla_{\underline{\mu}} \ \log\{p(\underline{x}_k|\underline{\mu})\} = \Sigma^{-1}\underline{x}_k - \Sigma^{-1}\underline{\mu} \qquad (3-7)$$

The ML estimator, $\hat{\underline{\mu}}$, therefore satisfies

$$\nabla_{\underline{\mu}} l(\hat{\underline{\mu}}) = \sum_{k=1}^{n}\{\Sigma^{-1}\underline{x}_k - \Sigma^{-1}\hat{\underline{\mu}}\} = \underline{0} \qquad (3-8)$$

Since we assumed Σ is invertible[3], (3-8) becomes

$$\sum_{k=1}^{n}\underline{x}_k = n\hat{\underline{\mu}} \qquad (3-9)$$

or

$$\hat{\underline{\mu}} = \frac{1}{n}\sum_{k=1}^{n}\underline{x}_k \qquad (3-10)$$

Thus, the ML estimate for the mean of a Gaussian distribution is the *sample mean*. This result, while rigorously developed, is intuitively appealing. ■

(ב) שערוך מטריצת הקווריאנס (בהינתן וקטורי התוחלות) :

Without proof, we show the ML estimator for the covariance matrix

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(\underline{x}_k-\hat{\underline{\mu}})(\underline{x}_k-\hat{\underline{\mu}})^T \qquad (3-11)$$

The ML estimator for Σ is biased, since $E\{\hat{\Sigma}\} \neq \Sigma$. However, the following is an unbiased estimator:

$$\hat{\Sigma}_u = \frac{1}{n-1}\sum_{k=1}^{n}(\underline{x}_k-\hat{\underline{\mu}})(\underline{x}_k-\hat{\underline{\mu}})^T \qquad (3-12)$$

■

שאלה 5 – שערוך התוחלת והקווריאנס של הפילוג הגאוסי

בדרך כלל, יש לשערך הן את התוחלת והן את הקוורואנס של הפילוג הגאוסי. נתייחס למקרה של מאפיין יחיד

In the general (and more typical) multivariate normal case, neither the mean $\boldsymbol{\mu}$ nor the covariance matrix Σ is known. Thus, these unknown parameters constitute the components of the parameter vector $\boldsymbol{\theta}$. Consider the univariate case with $\theta_1 = \mu$ and $\theta_2 = \sigma^2$. Here

$$\log p(x_k \mid \boldsymbol{\theta}) = -\tfrac{1}{2} \log 2\pi\theta_2 - \frac{1}{2\theta_2}(x_k - \theta_1)^2$$

and

$$\nabla_{\boldsymbol{\theta}} \log p(x_k \mid \boldsymbol{\theta}) = \begin{bmatrix} \dfrac{1}{\theta_2}(x_k - \theta_1) \\ -\dfrac{1}{2\theta_2} + \dfrac{(x_k - \theta_1)^2}{2\theta_2^2} \end{bmatrix}.$$

Then, Eq. (5) leads to the conditions

$$\sum_{k=1}^{n} \frac{1}{\hat{\theta}_2}(x_k - \hat{\theta}_1) = 0$$

and

$$-\sum_{k=1}^{n} \frac{1}{\hat{\theta}_2} + \sum_{k=1}^{n} \frac{(x_k - \hat{\theta}_1)^2}{\hat{\theta}_2^2} = 0,$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the maximum likelihood estimates for θ_1 and θ_2, respectively. By substituting $\hat{\mu} = \hat{\theta}_1$, $\hat{\sigma}^2 = \hat{\theta}_2$ and doing a little rearranging, we obtain the following maximum likelihood estimates for μ and σ^2:

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k \tag{7}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \hat{\mu})^2. \tag{8}$$

While the analysis of the multivariate case is basically very similar, considerably more manipulation is involved. The well known* result is that the maximum likelihood estimates for $\boldsymbol{\mu}$ and Σ are given by

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \tag{9}$$

and

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} (\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^t. \tag{10}$$

Thus, once again we find that the maximum likelihood estimate for the mean vector is the sample mean. The maximum likelihood estimate for the covariance matrix is the arithmetic average of the n matrices $(\mathbf{x}_k - \hat{\boldsymbol{\mu}}) \times (\mathbf{x}_k - \hat{\boldsymbol{\mu}})^t$. Since the true covariance matrix is the expected value of the matrix $(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t$, this is also a very satisfying result.