Pattern Recognition Assignment 1

Doris Yang

October 23, 2014

1 Dishonest Gambler

$$\begin{aligned}
Pr(6|A) &= 1/6 \\
Pr(6|B) &= 0.8 \\
Pr(6|C) &= 0.04 \\
Pr(B|6) &= \frac{P(B) \times P(6|B)}{P(A) \times P(6|A) + P(B) \times P(6|B) + P(C) \times P(6|C)} \\
&= \frac{0.2 \times 0.8}{0.6 \times 1/6 + 0.2 \times 0.8 + 0.2 \times 0.04} \\
&= 0.597
\end{aligned}$$

2 Three-class Classification

classes ω_1 ω_2 ω_3 are equally probable with class-conditional $p(x|\omega_i) = N(\mu_i, \Sigma_i)$ with

$$\mu_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 2 \\ -2 \end{pmatrix} \quad \mu_3 = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

and identical covariance matrices $\Sigma_1 = \Sigma_2 = \Sigma_3 = \sigma^2 I$ where $\sigma^2 = 0.25$

2.1 Discriminant Functions

In general, the discriminant function is

$$g_i(x) = -\tfrac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + lnP(\omega_i) + c_i$$

In this question, since all covariance matrices are the same, thus

$$\begin{aligned}
g_i(x) &= \omega_i^T x + \omega_{i0} \\
\omega_i &= \Sigma^{-1}\mu_i \\
\omega_{i0} &= lnP(\omega_i) - \tfrac{1}{2}\mu_i^T \Sigma^{-1}\mu_i
\end{aligned}$$

Because ω_1, ω_2, ω_3 are equally probable, $P(\omega_1) = P(\omega_2) = P(\omega_3) = \frac{1}{3}$

$$\omega_1 = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \end{bmatrix}$$
$$\omega_{i0} = ln\frac{1}{3} - \frac{1}{2} \begin{bmatrix} 2 \\ 2 \end{bmatrix}^T \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$
$$= -1.099 - 16 = -17.099$$
$$\therefore g_1(x) = \begin{bmatrix} 8 \\ 8 \end{bmatrix}^T x - 17.099$$

$$\omega_1 = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ -8 \end{bmatrix}$$
$$\omega_{i0} = ln\frac{1}{3} - \frac{1}{2} \begin{bmatrix} 2 \\ -2 \end{bmatrix}^T \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$
$$= -1.099 - 16 = -17.099$$
$$\therefore g_2(x) = \begin{bmatrix} 8 \\ -8 \end{bmatrix}^T x - 17.099$$

$$\omega_1 = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -8 \\ 0 \end{bmatrix}$$
$$\omega_{i0} = ln\frac{1}{3} - \frac{1}{2} \begin{bmatrix} -2 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$
$$= -1.099 - 8 = -9.099$$
$$\therefore g_3(x) = \begin{bmatrix} -8 \\ 0 \end{bmatrix}^T x - 9.099$$

2.2 Decision Regions and Boundaries

2.2.1 Decision Boundaries

Let $g_1(x) = g_2(x)$:

$$l_1 : \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T x = 0$$

Let $g_2(x) = g_3(x)$:

$$l_2 : \begin{bmatrix} 2 \\ -1 \end{bmatrix}^T x - 1 = 0$$

Let $g_1(x) = g_3(x)$:

$$l_3 : \begin{bmatrix} 2 \\ 1 \end{bmatrix}^T x - 1 = 0$$

2.2.2 Decision Region

- class ω_1: the region above l_1 and l_3
- class ω_2: the region below l_1 and l_2
- class ω_3: the region above $l2$ and below l_3

2.3 Plots

Figure 1 shows the distribution and corresponding decision boundaries for the three classes.

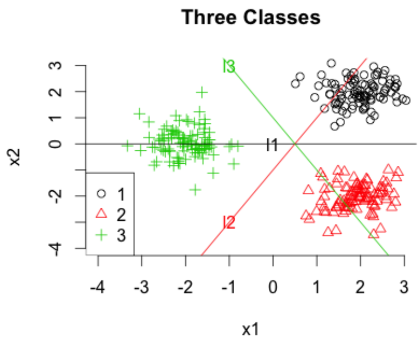

Figure 1: Decision Boundaries for Samples from Three Classes

2.4 MLE

For multivariate Gaussian distribution:

$$p(\boldsymbol{x_k};\boldsymbol{\mu}) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(\boldsymbol{x_k}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x_k}-\boldsymbol{\mu})\right)$$

Construct the likelihood function:

$$L(\boldsymbol{\mu_i}) = L(\boldsymbol{\Sigma_i}) = N \cdot ln\frac{1}{(2\pi)^{l/2}|\Sigma_i|^{1/2}} - \frac{1}{2}\sum_{k=1}^{N}(\boldsymbol{x_k}-\boldsymbol{\mu_i})^T\Sigma^{-1}(\boldsymbol{x_k}-\boldsymbol{\mu_i})$$

To calculate the maximum likelihood estimate, let $\dfrac{\partial L(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = 0$ and $\dfrac{\partial L(\Sigma)}{\partial \Sigma} = 0$

$$\frac{\partial L(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} \equiv \begin{bmatrix} \frac{\partial L}{\partial \boldsymbol{\mu_1}} \\ \frac{\partial L}{\partial \boldsymbol{\mu_2}} \\ \vdots \\ \frac{\partial L}{\partial \boldsymbol{\mu_l}} \end{bmatrix} = \sum_{k=1}^{N}(\boldsymbol{x_k}-\boldsymbol{\mu})\Sigma^{-1} = 0$$

$$\therefore \hat{\boldsymbol{\mu}} = \frac{\sum_{k=1}^{N}\boldsymbol{x_k}}{N}$$

$$\frac{\partial L(\Sigma)}{\partial \Sigma} = 0$$
$$\therefore \hat{\Sigma} = \frac{1}{N}\sum_{k=1}^{N}(\boldsymbol{x_k} - \boldsymbol{\hat{\mu}})(\boldsymbol{x_k} - \boldsymbol{\hat{\mu}})^T$$

Thus,

- $\hat{\mu_1} = \begin{bmatrix} 1.987 \\ 1.971 \end{bmatrix} \qquad \hat{\Sigma_1} = \begin{bmatrix} 0.264 & 0.017 \\ 0.017 & 0.228 \end{bmatrix}$
- $\hat{\mu_2} = \begin{bmatrix} 1.909 \\ -2.042 \end{bmatrix} \hat{\Sigma_2} = \begin{bmatrix} 0.237 & 0.051 \\ 0.051 & 0.258 \end{bmatrix}$
- $\hat{\mu_3} = \begin{bmatrix} -1.996 \\ 0.045 \end{bmatrix} \hat{\Sigma_3} = \begin{bmatrix} 0.227 & -0.016 \\ -0.016 & 0.291 \end{bmatrix}$

2.5 Decision Boundaries

In this case, the covariance matrices are arbitrary. Thus, the resulting discriminant function are quadratic:

$$\begin{aligned} g_i(x) &= \boldsymbol{x^T W_i x} + \boldsymbol{w_i^T x} + \omega_{i0} \\ \boldsymbol{W_i} &= \frac{1}{2}\Sigma_i^{-1} \\ \boldsymbol{w_i} &= \Sigma_i^{-1}\boldsymbol{\mu_i} \\ \omega_{i0} &= -\frac{1}{2}\boldsymbol{\mu_i}^T\Sigma_i^{-1}\boldsymbol{\mu_i} - \frac{1}{2}ln|\Sigma_i| + lnP(\omega_i) \end{aligned}$$

$$\begin{aligned} g_1'(x) &= \boldsymbol{x}^T \begin{bmatrix} 1.902 & -0.146 \\ -0.146 & 2.208 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 6.984 \\ 8.124 \end{bmatrix}^T \boldsymbol{x} - 14.945 \\ g_2'(x) &= \boldsymbol{x}^T \begin{bmatrix} 2.201 & -0.436 \\ -0.436 & 2.026 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 10.188 \\ -9.938 \end{bmatrix}^T \boldsymbol{x} - 19.874 \\ g_3'(x) &= \boldsymbol{x}^T \begin{bmatrix} 2.214 & 0.123 \\ 0.123 & 1.728 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} -8.828 \\ -0.335 \end{bmatrix}^T \boldsymbol{x} - 8.803 \end{aligned}$$

Let $g_1'(x) = g_2'(x)$:

$$l_1' : \boldsymbol{x}^T \begin{bmatrix} -0.300 & 0.290 \\ 0.290 & 0.182 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} -3.203 \\ 18.062 \end{bmatrix}^T \boldsymbol{x} + 4.929 = 0$$

Let $g_2'(x) = g_3'(x)$:

$$l_2' : \boldsymbol{x}^T \begin{bmatrix} -0.013 & -0.559 \\ -0.559 & 0.298 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 19.016 \\ -9.603 \end{bmatrix}^T \boldsymbol{x} - 11.070 = 0$$

Let $g'_1(x) = g'_3(x)$:

$$l'_3 : \boldsymbol{x}^T \begin{bmatrix} -0.312 & -0.268 \\ -0.268 & 0.480 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 15.813 \\ 8.459 \end{bmatrix}^T \boldsymbol{x} - 6.142 = 0$$

Figure 2 shows the decision boundaries with the parameters estimated by MLE.

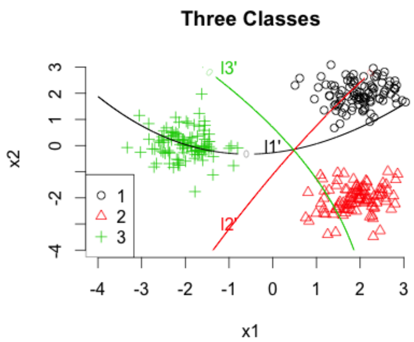

Figure 2: Decision Boundaries with the maximum likelihood estimates

2.6 Another Sample

Regenerate a new set of samples and estimate $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma_i}$ with MLE. The following boundaries can be obtained. And corresponding plot is showed in Figure 3

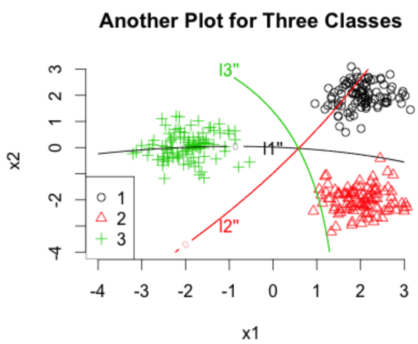

Figure 3: Another set of data sample with corresponding decison boundaries

$$l_1'' : \boldsymbol{x}^T \begin{bmatrix} 0.557 & -0.324 \\ -0.324 & 0.174 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 1.036 \\ 16.155 \end{bmatrix}^T \boldsymbol{x} - 0.294 = 0$$

$$l_2'' : \boldsymbol{x}^T \begin{bmatrix} 0.128 & 0.230 \\ 0.230 & 0.076 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 15.522 \\ -9.542 \end{bmatrix}^T \boldsymbol{x} - 9.387 = 0$$

$$l_3'' : \boldsymbol{x}^T \begin{bmatrix} 0.686 & -0.932 \\ -0.932 & 0.249 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 16.558 \\ 6.614 \end{bmatrix}^T \boldsymbol{x} - 9.680 = 0$$

Explanation Decision boundaries obtained from two groups of data sample are different. But they also share some similarities. The differences are caused by the different data samples used to estimate the parameters. While the similar parts are because two groups of data samples are generated by the same distribution.