Open your mind. LUT.
Lappeenranta University of Technology

LUT Machine Vision and Pattern Recognition 2015-11-30

BM40A0700 Pattern Recognition

Lasse Lensu

Exercise 5 solutions: Bayesian classification

1. Effect of unequal variances (2 *bonus* points): The task was related to two normal probability distributions
$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right] \tag{1}$$
with **equal** a priori probabilities $P(\omega_1) = P(\omega_2)$ and with **unequal** standard deviations (STDs) $\sigma_1 = 1$ and $\sigma_2 = 2$. The purpose was to determine a minimum-error classifier for the setting where $\mu_1 = 3$ and $\mu_2 = 6$.

   Discriminant function $g_i(x)$ is the a posteriori probability $P(\omega_i|x)$. Thus, using Bayes formula
$$g_i(x) = P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$$

   After taking natural logarithm, standard logarithm rules can be used to refine the result
$$\begin{aligned}
g_i'(x) = \ln P(\omega_i|x) &= \ln\left[\frac{p(x|\omega_i)P(\omega_i)}{p(x)}\right] = \ln p(x|\omega_i) + \ln P(\omega_i) - \ln p(x)\\
&= \ln\left\{\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right]\right\} + \ln P(\omega_i) - \ln p(x)\\
&= \ln\left\{\exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right]\right\} - \ln\left(\sqrt{2\pi}\sigma_i\right) + \ln P(\omega_i) - \ln p(x)\\
&= -\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2 - \ln\left(\sqrt{2\pi}\sigma_i\right) + \ln P(\omega_i) - \ln p(x)
\end{aligned}$$

   When the discriminant functions are set equal and unnecessary terms are removed:
$$\begin{aligned}
g_1'(x) &= g_2'(x)\\
-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \ln\left(\sqrt{2\pi}\sigma_1\right) &= -\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 + \ln\left(\sqrt{2\pi}\sigma_2\right)\\
-\frac{1}{2\sigma_1^2}\left(x^2 - 2\mu_1 x + \mu_1^2\right) + \ln\left(\sqrt{2\pi}\sigma_1\right) &= -\frac{1}{2\sigma_2^2}\left(x^2 - 2\mu_2 x + \mu_2^2\right) + \ln\left(\sqrt{2\pi}\sigma_2\right)\\
\frac{1}{\sigma_2^2}x^2 - \frac{1}{\sigma_1^2}x^2 + \frac{2\mu_1 x}{\sigma_1^2} - \frac{2\mu_2 x}{\sigma_2^2} + \frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + 2\ln\left(\frac{\sigma_2}{\sigma_1}\right) &= 0\\
\left(\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2\sigma_2^2}\right)x^2 + 2\left(\frac{\mu_1\sigma_2^2-\mu_2\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)x + \frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + 2\ln\left(\frac{\sigma_2}{\sigma_1}\right) &= 0
\end{aligned}$$

   This is a polynomial equation of quadratic form, and we can use the given values, $\mu_1 = 3$, $\mu_2 = 6$,

PO BOX 20, FI-53851 LAPPEENRANTA, FINLAND
E-mail: lasse.lensu@lut.fi

$\sigma_1 = 1$ and $\sigma_2 = 2$, to get the coefficients a, b, c for the common solution for such an equation

$$\begin{aligned} x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad \text{where} \\ a &= \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 \sigma_2^2} = -\frac{3}{4} \\ b &= 2\left(\frac{\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2}{\sigma_1^2 \sigma_2^2}\right) = 3 \\ c &= \frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} + 2 \ln\left(\frac{\sigma_2}{\sigma_1}\right) \approx 1.386 \\ x &\approx \begin{cases} -0.418 \\ 4.42 \end{cases} \end{aligned}$$

An example visualisation of the case is given in Fig. 1.

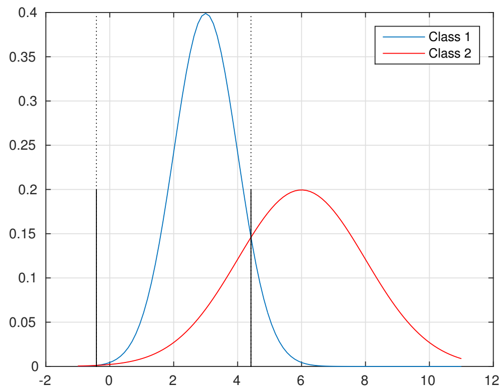

Figure 1: The classes, and the decision boundaries determined by using the symbolically and manually determined solutions.

2. Classification in two dimensions with unequal covariances (2 points): The following three class classification problem was considered. Each class is normally distributed with unequal covariance matrix and equal a priori probabilities.

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{\mu}_3 = \begin{pmatrix} 11 \\ -2 \end{pmatrix} \quad \boldsymbol{\Sigma}_3 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$P(\omega_1) = P(\omega_2) = P(\omega_3) \quad .$$

$$|\mathbf{\Sigma}_1| = 1 \quad {\mathbf{\Sigma}_1}^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$|\mathbf{\Sigma}_2| = 4 \quad {\mathbf{\Sigma}_2}^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$|\mathbf{\Sigma}_3| = 1 \quad {\mathbf{\Sigma}_3}^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

Discriminant function is

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_i)^T {\mathbf{\Sigma}_i}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i) - \left(\frac{d}{2}\right)\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Sigma}_i| \quad ,$$

where d is dimensionality of samples, in this case $d = 2$. Term $\left(\frac{d}{2}\right)\ln(2\pi)$ is a constant term and can be ignored.

Discriminant function for the first class is

$$\begin{aligned}
g_1(\boldsymbol{x}) &= -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)^T {\Sigma_1}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) - \frac{1}{2}\ln|\mathbf{\Sigma}_1| \\
&= -\frac{1}{2}\begin{pmatrix} x_1 - 3 \\ x_2 - 6 \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} x_1 - 3 \\ x_2 - 6 \end{pmatrix} - \frac{1}{2}\ln 1 \\
&= -\frac{1}{2}\begin{pmatrix} x_1 - 3 & x_2 - 6 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} x_1 - 3 \\ x_2 - 6 \end{pmatrix} \\
&= -\frac{1}{2}\begin{pmatrix} 2x_1 - 6 & \frac{1}{2}x_2 - 3 \end{pmatrix}\begin{pmatrix} x_1 - 3 \\ x_2 - 6 \end{pmatrix} \\
&= -\frac{1}{2}\left[(2x_1 - 6)(x_1 - 3) + \left(\frac{1}{2}x_2 - 3\right)(x_2 - 6)\right] \\
&= -\frac{1}{2}\left[2{x_1}^2 - 6x_1 - 6x_1 + 18 + \frac{1}{2}{x_2}^2 - 3x_2 - 3x_2 + 18\right] \\
&= -\frac{1}{2}\left[2{x_1}^2 - 12x_1 + \frac{1}{2}{x_2}^2 - 6x_2 + 36\right] \\
&= -{x_1}^2 + 6x_1 - \frac{1}{4}{x_2}^2 + 3x_2 - 18 \quad .
\end{aligned}$$

Discriminant function for the second class is

$$\begin{aligned}
g_2(\boldsymbol{x}) &= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^T{\Sigma_2}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_2| \\
&= -\frac{1}{2}\begin{pmatrix} x_1-3 \\ x_2+2 \end{pmatrix}^T \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x_1-3 \\ x_2+2 \end{pmatrix} - \frac{1}{2}\ln 4 \\
&= -\frac{1}{2}\begin{pmatrix} x_1-3 & x_2+2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x_1-3 \\ x_2+2 \end{pmatrix} - \ln 2 \\
&= -\frac{1}{2}\begin{pmatrix} \frac{1}{2}x_1-\frac{3}{2} & \frac{1}{2}x_2+1 \end{pmatrix} \begin{pmatrix} x_1-3 \\ x_2+2 \end{pmatrix} - \ln 2 \\
&= -\frac{1}{2}\left[\left(\frac{1}{2}x_1-\frac{3}{2}\right)(x_1-3) + \left(\frac{1}{2}x_2+1\right)(x_2+2)\right] - \ln 2 \\
&= -\frac{1}{2}\left[\frac{1}{2}{x_1}^2 - \frac{3}{2}x_1 - \frac{3}{2}x_1 + \frac{9}{2} + \frac{1}{2}{x_2}^2 + x_2 + x_2 + 2\right] - \ln 2 \\
&= -\frac{1}{2}\left[\frac{1}{2}{x_1}^2 - 3x_1 + \frac{1}{2}{x_2}^2 + 2x_2 + \frac{13}{2}\right] - \ln 2 \\
&= -\frac{1}{4}{x_1}^2 + \frac{3}{2}x_1 - \frac{1}{4}{x_2}^2 - x_2 - \frac{13}{4} - \ln 2 \quad .
\end{aligned}$$

Discriminant function for the third class is

$$\begin{aligned}
g_3(\boldsymbol{x}) &= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_3)^T{\Sigma_3}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_3) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_3| \\
&= -\frac{1}{2}\begin{pmatrix} x_1-11 \\ x_2+2 \end{pmatrix}^T \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1-11 \\ x_2+2 \end{pmatrix} - \frac{1}{2}\ln 1 \\
&= -\frac{1}{2}\begin{pmatrix} x_1-11 & x_2+2 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1-11 \\ x_2+2 \end{pmatrix} \\
&= -\frac{1}{2}\begin{pmatrix} \frac{1}{2}x_1-\frac{11}{2} & 2x_2+4 \end{pmatrix} \begin{pmatrix} x_1-11 \\ x_2+2 \end{pmatrix} \\
&= -\frac{1}{2}\left[\left(\frac{1}{2}x_1-\frac{11}{2}\right)(x_1-11) + (2x_2+4)(x_2+2)\right] \\
&= -\frac{1}{2}\left[\frac{1}{2}{x_1}^2 - \frac{11}{2}x_1 - \frac{11}{2}x_1 + \frac{121}{2} + 2{x_2}^2 + 4x_2 + 4x_2 + 8\right] \\
&= -\frac{1}{2}\left[\frac{1}{2}{x_1}^2 - 11x_1 + 2{x_2}^2 + 8x_2 + \frac{137}{2}\right] \\
&= -\frac{1}{4}{x_1}^2 + \frac{11}{2}x_1 - {x_2}^2 - 4x_2 - \frac{137}{4} \quad .
\end{aligned}$$

Decision boundary between the first and second class is defined by equation

$$\begin{aligned}
g_1(\boldsymbol{x}) &= g_2(\boldsymbol{x}) \\
-{x_1}^2 + 6x_1 - \frac{1}{4}{x_2}^2 + 3x_2 - 18 &= -\frac{1}{4}{x_1}^2 + \frac{3}{2}x_1 - \frac{1}{4}{x_2}^2 - x_2 - \frac{13}{4} - \ln 2 \\
-{x_1}^2 + \frac{1}{4}{x_1}^2 - \frac{1}{4}{x_2}^2 + \frac{1}{4}{x_2}^2 &= -6x_1 - 3x_2 + \frac{3}{2}x_1 - x_2 + 18 - \frac{13}{4} - \ln 2 \\
-\frac{3}{4}{x_1}^2 &= -\frac{9}{2}x_1 - 4x_2 + 18 - \frac{13}{4} - \ln 2 \\
4x_2 &= \frac{3}{4}{x_1}^2 - \frac{9}{2}x_1 + 18 - \frac{13}{4} - \ln 2 \\
x_2 &= \frac{3}{16}{x_1}^2 - \frac{9}{8}x_1 + \frac{59}{16} - \frac{\ln 2}{4} \quad .
\end{aligned}$$

Decision boundary between the first and second class is parabola which opens to the direction of positive x_2-axis.

Decision boundary between the second and third class is defined by equation

$$\begin{aligned}
g_2(\boldsymbol{x}) &= g_3(\boldsymbol{x}) \\
-\frac{1}{4}{x_1}^2 + \frac{3}{2}x_1 - \frac{1}{4}{x_2}^2 - x_2 - \frac{13}{4} - \ln 2 &= -\frac{1}{4}{x_1}^2 + \frac{11}{2}x_1 - {x_2}^2 - 4x_2 - \frac{137}{4} \\
\frac{3}{2}x_1 - \frac{11}{2}x_1 &= \frac{1}{4}{x_2}^2 - {x_2}^2 + x_2 - 4x_2 + \frac{13}{4} + \ln 2 - \frac{137}{4} \\
-\frac{8}{2}x_1 &= -\frac{3}{4}{x_2}^2 - 3x_2 - \frac{124}{4} + \ln 2 \\
x_1 &= \frac{3}{16}{x_2}^2 + \frac{3}{4}x_2 + \frac{31}{4} - \frac{\ln 2}{4} \quad .
\end{aligned}$$

Decision boundary between the second and third class is parabola which opens to the direction of positive x_1-axis.

Decision boundary between the first and third class is defined by equation

$$\begin{aligned}
g_1(\boldsymbol{x}) &= g_3(\boldsymbol{x}) \\
-{x_1}^2 + 6x_1 - \frac{1}{4}{x_2}^2 + 3x_2 - 18 &= -\frac{1}{4}{x_1}^2 + \frac{11}{2}x_1 - {x_2}^2 - 4x_2 - \frac{137}{4} \\
-{x_1}^2 + \frac{1}{4}{x_1}^2 + 6x_1 - \frac{11}{2}x_1 &= \frac{1}{4}{x_2}^2 - {x_2}^2 - 3x_2 - 4x_2 + 18 - \frac{137}{4} \\
-\frac{3}{4}{x_1}^2 + \frac{1}{2}x_1 &= -\frac{3}{4}{x_2}^2 - 7x_2 + 18 - \frac{137}{4} \\
3{x_1}^2 - 2x_1 &= 3{x_2}^2 + 28x_2 + 65 \quad .
\end{aligned}$$

We can write this equation as a second degree polynomial of x_2:

$$3{x_2}^2 + 28x_2 - 3{x_1}^2 + 2x_1 + 65 = 0$$

From this equation x_2 can be solved using standard formula for second degree polynomials:

$$\begin{aligned}
x_2 &= \frac{-28 \pm \sqrt{28^2 - 4 \cdot 3 \cdot (-3{x_1}^2 + 2x_1 + 65)}}{2 \cdot 3} \\
&= \frac{-28 \pm \sqrt{784 + 36{x_1}^2 - 24x_1 - 780}}{6} \\
&= \frac{-28 \pm \sqrt{36{x_1}^2 - 24x_1 + 4}}{6} \\
&= \frac{-28 \pm \sqrt{(6x_1 - 2)^2}}{6} \\
&= \frac{-28 \pm (6x_1 - 2)}{6}
\end{aligned}$$

We have solutions

$$\begin{cases} x_{21} = \frac{-28+(6x_1-2)}{6} = x_1 - 5 \\ x_{21} = \frac{-28-(6x_1-2)}{6} = -x_1 - \frac{13}{3} \end{cases} .$$

The decision boundary between the second and third class consist of two divergent lines.

The class centers, classes and decision boundaries are illustrated in Fig. [**?**].

3. Statistical classifier with normal distribution (2 points): not published.

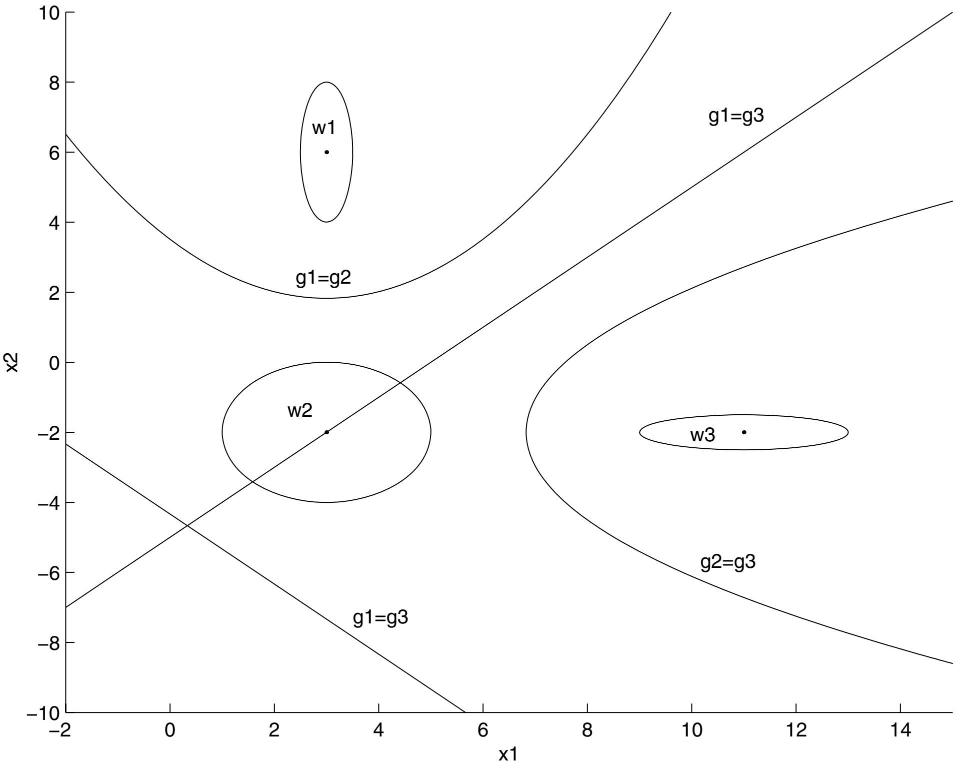

Figure 2: Class centers, classes and decision boundaries in the case of three given classes.