Introduction to Machine Learning - HW2 - Q4

Professors: Abolghasemi & Arabi
Student: Mohamad Mahdi Samadi
Student ID: 810101465

Given a dataset with two 2D classes C_1 and C_2, each coming from a Gaussian distribution and prior of $P(C_1) = 0.6$ and $P(C_2) = 0.4$, find the decision boundary between the two classes.

$$C_1: N(\mu_1, \Sigma_1), \quad \mu_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C_2: N(\mu_2, \Sigma_2), \quad \mu_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

PDF of a gaussian distribution in d dimensions is:

$$P(x|C_i) = (2\pi)^{\frac{-d}{2}} |\Sigma_i|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}$$

Where:

- d is the number of dimensions.
- μ_i is the mean vector of class C_i
- Σ_i is the covariance matrix of class C_i
- $|\Sigma_i|$ is the determinant of the covariance matrix.
- Σ_i^{-1} is the inverse of the covariance matrix.
- $(x - \mu_i)^T \Sigma^{-1} (x - \mu_i)$ is the Mahalanobis distance.

In our case $d = 2$, so:

$$P(x|C_1) = (2\pi)^{-1} |\Sigma_1|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_1)^T \Sigma_1^{-1} (x-\mu_1)}$$

$$P(x|C_2) = (2\pi)^{-1} |\Sigma_2|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_2)^T \Sigma_2^{-1} (x-\mu_2)}$$

To find the decision boundary we'll have to solve this equation:

$g_1(x) = g_2(x)$ where $g_i(x) = P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)} \rightarrow$

$$\frac{P(x|C_1)P(C_1)}{P(x)} = \frac{P(x|C_2)P(C_2)}{P(x)}$$

Knowing the prior probabilities and the fact that $P(x)$ is the same for both of them, we'll continue to solve the equation.

$$0.6 \times P(x|C_1) = 0.4 \times P(x|C_2)$$

$$0.6 \times (2\pi)^{-1}|\Sigma_1|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_1)^T\Sigma_2^{-1}(x-\mu_1)} = 0.4 \times (2\pi)^{-1}|\Sigma_2|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}$$

$$3 \times |\Sigma_1|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_1)^T\Sigma_2^{-1}(x-\mu_1)} = 2 \times |\Sigma_2|^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}$$

$$3 \times 4^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)} = 2 \times 6^{\frac{-1}{2}} e^{\frac{-1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}$$

$$(\frac{3}{2})^{\frac{3}{2}} \times e^{\frac{-1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)} = e^{\frac{-1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}$$

Taking a logarithm leads us to:

$$\frac{-1}{2}(x - \mu_1)^T\Sigma_1^{-1}(x - \mu_1) + \frac{3}{2}ln(\frac{3}{2}) = \frac{-1}{2}(x - \mu_2)^T\Sigma_2^{-1}(x - \mu_2)$$

x is the vector of input $\{x_1, x_2\}$

$$(x - \mu_1) = \begin{bmatrix} x_1-2 \\ x_2-3 \end{bmatrix} \rightarrow (x - \mu_1)^T = [x_1 - 2, x_2 - 3]$$

$$\Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \Sigma_1^{-1} = \frac{1}{4}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

$$(x - \mu_2) = \begin{bmatrix} x_1-5 \\ x_2-1 \end{bmatrix} \rightarrow (x - \mu_2)^T = [x_1 - 5, x_2 - 1]$$

$$\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \rightarrow \Sigma_2^{-1} = \frac{1}{6}\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\frac{-1}{2}(x - \mu_1)^T\Sigma_1^{-1}(x - \mu_1) = \frac{-1}{8}[x_1 - 2, x_2 - 3] \times \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \times \begin{bmatrix} x_1-2 \\ x_2-3 \end{bmatrix}$$

$$\frac{-1}{8}(x_1 - 2)^2 - \frac{1}{2}(x_2 - 3)^2$$

$$\frac{-1}{2}(x - \mu_2)^T\Sigma_2^{-1}(x - \mu_2) = \frac{-1}{12}[x_1 - 5, x_2 - 1] \times \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \times \begin{bmatrix} x_1-5 \\ x_2-1 \end{bmatrix}$$

$$= \frac{-1}{4}(x_1 - 5)^2 - \frac{1}{6}(x_2 - 1)^2$$

$$\frac{1}{8}(x_1 - 2)^2 + \frac{1}{2}(x_2 - 3)^2 = \frac{1}{4}(x_1 - 5)^2 + \frac{1}{6}(x_2 - 1)^2 + \frac{3}{2}log(\frac{3}{2})$$

By multiplying both sides at 24 we'll have:

$$3(x_1 - 2)^2 + 12(x_2 - 3)^2 = 6(x_1 - 5)^2 + 4(x_2 - 1)^2 + 36\, log(\frac{3}{2})$$

$$8x_2^2 - 3x_1^2 + 48x_1 - 64x_2 - 34 - 36\, ln(\frac{3}{2}) = 0$$

$$8(x_2 - 4)^2 - 3(x_1 - 8)^2 + 30 + 36\, ln(\frac{3}{2}) = 0$$

The final equation shows the decision hyperplane in a 3D space (the input vector x and output y) or the decision curve in a 2D space (just the input vector x).