Machine Learning - Theoretical exercise 3

Téo Bouvard

January 26, 2020

Problem 1

a) Assuming an gaussian distribution $X \sim \mathcal{N}(\mu, \Sigma)$ the maximum likelihood method states that for a set of measurements $\chi = \{x_1, \ldots, x_N\}$,

$$\mu = \frac{1}{N} \sum_{k=1}^{N} x_k \tag{1}$$

$$\Sigma = \frac{1}{N} \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{2}$$

We first estimate the mean vectors of the two distributions using (1)

$$\mu_1 = \frac{1}{4} \left(\begin{bmatrix} 2 \\ 6 \end{bmatrix} + \begin{bmatrix} 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 \\ 8 \end{bmatrix} + \begin{bmatrix} 4 \\ 6 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 12 \\ 24 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

$$\mu_2 = \frac{1}{4} \left(\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2.7 \\ -4 \end{bmatrix} + \begin{bmatrix} 3.3 \\ 0 \end{bmatrix} + \begin{bmatrix} 5 \\ -2 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 12 \\ -8 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

We use the estimated mean vectors to compute the covariance matrices according to (2)

$$\Sigma_1 = \frac{1}{4} \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$

$$\Sigma_2 = \frac{1}{4} \left(\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0.09 & 0.6 \\ 0.6 & 4 \end{bmatrix} + \begin{bmatrix} 0.09 & 0.6 \\ 0.6 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \right) = \frac{1}{4} \begin{bmatrix} 8.18 & 1.2 \\ 1.2 & 8.18 \end{bmatrix} = \begin{bmatrix} 2.045 & 0.3 \\ 0.3 & 2 \end{bmatrix}$$

b) We use the log discriminant function to compute the decision boundary. Let $x = (x_1\, x_2)^T$ be on the decision boundary between the two distributions $\implies g_1(x) = g_2(x)$

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

We use the following properties of the covariances matrices Σ_1 and Σ_2 to simplify this equation.

$$|\Sigma_1| = 1 \implies \frac{1}{2} \ln |\Sigma_1| = 0$$

$$|\Sigma_2| = 4 \implies \frac{1}{2} \ln |\Sigma_1| = \ln 2$$

We then compute the two remaining terms independently.

$$\frac{1}{2}(x - \mu_1)^T \Sigma_1^{-1}(x - \mu_1) = \frac{1}{2} \begin{bmatrix} x_1 - 3 & x_2 - 6 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 - 3 \\ x_2 - 6 \end{bmatrix}$$

$$= \frac{1}{2} \left((2(x_1 - 3)^2 + \frac{1}{2}(x_2 - 6)^2 \right)$$

1

And

$$\frac{1}{2}(x - \mu_2)^T \Sigma_2^{-1}(x - \mu_2) = \frac{1}{2}\begin{bmatrix} x_1 - 3 & x_2 + 2 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & -\frac{3}{40} \\ -\frac{3}{40} & \frac{409}{800} \end{bmatrix}\begin{bmatrix} x_1 - 3 \\ x_2 + 2 \end{bmatrix}$$

$$= \frac{1}{2}\left(\frac{1}{2}(x_1 - 3)^2 - \frac{3}{20}(x_1 - 3)(x_2 + 2) + \frac{409}{800}(x_2 + 2)^2 \right)$$

Which gives us the final decision boundary equation

$$-\frac{3}{4}(x_1 - 3)^2 - \frac{1}{4}(x_2 - 6)^2 + \frac{409}{1600}(x_2 + 2)^2 - \frac{3}{40}(x_1 - 3)(x_2 + 2) + \ln 2 = 0$$

This equation describes a hyperbola whose upper part is slightly tilted towards the left. This is due to the samples of χ_2 not describing a circle, and thus orienting the decision border sideways.

c) In order to match more closely a parabolic decision boundary, one should gather more data samples in χ_1 and χ_2, which would smooth out irregularites in the samples.

Problem 2

Problem 3