Tutorial 4: Classification with Gaussians

1. Consider a pattern recognition problem with two classes A and B. Each class is modelled by a class-conditional Gaussian density. Class A is parameterised by mean $\mu_A = 4$ and variance $\sigma_A^2 = 1$; class B is parameterised by mean $\mu_B = 2$ and variance $\sigma_B^2 = 4$.

   (a) On the same graph sketch the probability density function for each class.

   (b) The following three test items are observed:

   $$\begin{aligned} x_1 &= 3 \\ x_2 &= 4 \\ x_3 &= 8 \end{aligned}$$

   Assume that the classes have equal prior probabilities. To which classes should these points be assigned?

   (c) You are told that the prior probability of class B is twice that of class A. To which classes would you now assign points x_1, x_2, x_3?

   (d) What are the benefits and drawbacks of using Gaussian probability density functions as a generative model for real world pattern recognition problems?

2. In a two-class pattern classification problem, with classes A and B, each class is modelled using a one-dimensional Gaussian probability density function:

   $$\begin{aligned} p(x|A) &= \mathcal{N}(x; \mu_A, \sigma_A^2) \\ p(x|B) &= \mathcal{N}(x; \mu_B, \sigma_B^2). \end{aligned}$$

   Assume the classes have equal prior probabilities, and that $\mu_A \neq \mu_B$ and $\sigma_A^2 \neq \sigma_B^2$.

   (a) Write down a suitable discriminant function for this problem.

   (b) Derive the quadratic equation in x that defines the decision boundary between the classes.

3. The notes stated without proof that the sample mean (μ_{ML}) and sample variance (σ^2_{ML}) are the maximum likelihood solutions for the parameters of a one-dimensional Gaussian. Consider the log likelihood of a Gaussian with mean μ and variance σ^2, given a set of N data points $\{x_1, \ldots, x_N\}$:

$$L = \ln p(\{x_1, \ldots, x_n\} | \mu, \sigma^2) = -\frac{1}{2} \sum_{n=1}^{N} \left(\frac{(x_n - \mu)^2}{\sigma^2} - \ln \sigma^2 - \ln(2\pi) \right)$$

$$= -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi) \, .$$

By maximising the log likelihood function with respect to μ show that the maximum likelihood estimate for the mean is the sample mean:

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n .$$

4. Consider a toy problem of two classes, C_1 and C_2, each of which has a normal distribution in a two-dimensional vector space, and assume there are some training samples for each class shown below:

$$C_1 \; : \; (1,2)^T, \; (2,0)^T, \; (2,4)^T, \; (3,2)^T$$
$$C_2 \; : \; (5,1)^T, \; (5,2)^T, \; (7,2)^T, \; (7,3)^T$$

(a) Estimate the mean vector $\hat{\boldsymbol{\mu}}_i$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_i$ for each class $i = 1, 2$ in terms of maximum likelihood. (It is advisable to do this at least by hand without using a calculator!)

(b) Using the parameters obtained above, sketch the contours of the normal distribution for each class.

(c) Find the eigen values and eigen vectors of each $\hat{\boldsymbol{\Sigma}}_i, \; i = 1, 2$, and discuss how they are related to the shape of the distribution. [non-examinable]