Probabilistic Unsupervised Learning
Exercise 2: Likelihood and Decision Theory

Submission deadline: Monday, November 21, 2022 at 10:00 a.m.

In the programming exercises, support is only provided for Matlab or Python source codes.

1. [4 *points*] Suppose $p(x|\lambda)$ is a one-dimensional Poisson probability function, $p(x|\lambda) = \exp(-\lambda)\frac{\lambda^x}{x!}$, with $x \in \mathbb{N}$. What is the maximum likelihood solution if N datapoints $x_1, ..., x_N$ are given ($x_n \in \mathbb{N}$)?

2. Suppose $\mathcal{N}(x; \mu, \sigma^2)$ denotes a Gaussian probability density function:

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) \tag{1}$$

Consider the function:

$$p(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{1}{2}\mathcal{N}(x; \mu_1, \sigma_1^2) + \frac{1}{2}\mathcal{N}(x; \mu_2, \sigma_2^2) \tag{2}$$

[1 *point*] Task A:
Show that $p(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is a probability density function (pdf).

[3 *points*] Task B:
Why is it difficult to find a maximum likelihood solution for $p(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$?
(Just answer the question and motivate the answer, do not try to find a solution.)

3. [3 *points*] Task A:
Generate $N = 100$ datapoints x_n from a one-dimensional Gaussian distribution. That is, draw 100 times values of a random variable x whose value distribution is described by a Gaussian pdf. Choose generating parameters $\mu = 20$ and $\sigma^2 = 4$ for the generation. Visualize the generated data as a scatter-plot and by using the binning formula of the lecture.

[2 *points*] Task B:
Compute the maximum likelihood parameters μ^* and $(\sigma^2)^*$ for the 100 datapoints generated in Task A.

[1 *point*] Task C:
Do you recover the generating parameters of Task A exactly? If yes, why? If no, why not?

[1 *point*] Task D:
What happens for smaller and for larger N?

4. [4 *points*] Task A:
Consider the optimal classification example with blueberries and oranges given in the lecture. Instead of 2 classes consider having K classes (K different fruits). How can you define decision regions R_k that result in an optimal classification of the fruits (in the minimal misclassification sense)? (Hint: it might be easier to compute p_{correct} instead of p_{mis}.)

   [1 *point*] Task B:
   Can a given R_k be the empty set? If yes, why? If no, why not?