Homework Assignment # 1
Due: Thursday, September 26, 2019, 11:59 p.m.
Total marks: 100

Question 1. [10 MARKS]

Let X be a random variable with outcome space $\Omega = \{a, b, c\}$ and $p(a) = 0.1, p(b) = 0.2$, and $p(c) = 0.7$. Let

$$f(x) = \begin{cases} 10 & \text{if } x = a \\ 5 & \text{if } x = b \\ 10/7 & \text{if } x = c \end{cases}$$

(a) [3 MARKS] What is $E[f(X)]$?

(b) [3 MARKS] What is $E[1/p(X)]$?

(c) [4 MARKS] For an arbitrary pmf p, what is $E[1/p(X)]$?

Question 2. [15 MARKS]

Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ be independent multivariate Gaussian random variables, with $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\mu}_i \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d \times d}$ for dimension $d \in \mathbb{N}$. Define $\mathbf{X} = a_1 \mathbf{X}_1 + a_2 \mathbf{X}_2 + \ldots + a_m \mathbf{X}_m$ as a convex combination, $a_i \geq 0$ and $\sum_{i=1}^m a_i = 1$.

(a) [5 MARKS] Write the expected value $E[\mathbf{X}]$ in terms of the givens $a_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$. Show all you steps. What is the dimension of $E[\mathbf{X}]$?

(b) [10 MARKS] Write the covariance $\text{Cov}[\mathbf{X}]$ in terms of the givens $a_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$. Show all you steps. What is the dimension of $\text{Cov}[\mathbf{X}]$? Briefly explain how the result for $\text{Cov}[\mathbf{X}]$ would be different if the variables $\mathbf{X}_1$ and $\mathbf{X}_2$ are not independent and have covariance $\text{Cov}[\mathbf{X}_1, \mathbf{X}_2] = \boldsymbol{\Lambda}$ for $\boldsymbol{\Lambda} \in \mathbb{R}^{d \times d}$.

Question 3. [15 MARKS]

This question involves some simple simulations, to better visualize random variables and get some intuition for sampling, which is a central theme in machine learning. Use the attached code called `simulate.py`. This code is a simple script for sampling and plotting with python; play with some of the parameters to see what it is doing. Calling `simulate.py` runs with default parameters; `simulate.py 1 100` simulates 100 samples from a 1d Gaussian. The generated plot is generically for 3 dimensions. If you call the function with 1d, then it simply plots the points on a line, but on a 3-dimensional plot. The maximum dimension that can be given to the script is 3.

Note that if you do not have matplotlib installed, you will have to install it.

(a) [5 MARKS] Run the code for 10, 100 and 1000 samples with dim=1 and $\sigma = 1.0$. Next run the code for 10, 100 and 1000 samples with dim=1 and $\sigma = 10.0$. What do you notice about the sample mean?

(b) [5 MARKS] The current covariance for dim=3 is

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

What does that mean about the multivariate Gaussian (i.e., the vector random variable composed of random variables X, Y and Z)?

(c) [5 MARKS] Change the covariance to

$$\Sigma = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

What happens?

Question 4. [30 MARKS]

Suppose that the number of accidents occurring daily in a certain plant has a Poisson distribution with an unknown mean λ. Based on previous experience in similar industrial plants, suppose that our initial feelings about the possible value of λ can be expressed by an exponential distribution with parameter $\theta = \frac{1}{2}$. That is, the prior density is

$$p(\lambda) = \theta e^{-\theta \lambda}$$

where $\lambda \in [0, \infty)$.

(a) [5 MARKS] Before observing any data (any reported accidents), what is the most likely value for λ?

(b) [5 MARKS] Now imagine there are 79 accidents over 9 days. Determine the maximum likelihood estimate of λ.

(c) [5 MARKS] Again imagine there are 79 accidents over 9 days. Determine the maximum a posteriori (MAP) estimate of λ.

(d) [5 MARKS] Imagine you now want to predict the number of accidents for tomorrow. How can you use the maximum likelihood estimate computed above? What about the MAP estimate? What would they predict?

(e) [5 MARKS] For the MAP estimate, what is the purpose of the prior once we observe this data?

(f) [5 MARKS] Imagine that now new safety measures have been put in place and you believe that the number of accidents per day should sharply decrease. How might you change θ to better reflect this new belief about the number of accidents? *Hint:* Look at the plots of some exponential distributions to better understand the prior chosen on λ.

Question 5. [30 MARKS]

Imagine that you would like to predict if your favorite table will be free at your favorite restaurant. The only additional piece of information you can collect, however, is if it is sunny or not sunny. You collect paired samples from visit of the form (is sunny, is table free), where it is either sunny (1) or not sunny (0) and the table is either free (1) or not free(0).

(a) [10 MARKS] How can this be formulated as a maximum likelihood problem?

(b) [10 MARKS] Assume you have collected data for the last 10 days and computed the maximum likelihood solution to the problem formulated in (a). If it is sunny today, how would you predict if your table will be free?

(c) [10 MARKS] Imagine now that you could further gather information about if it is morning, afternoon, or evening. How does this change the maximum likelihood problem?

Bonus.   [20 MARKS]

(d) [10 MARKS] Using a computer, generate 1000 samples from a d-dimensional multivariate Gaussian with mean $\mathbf{0}$ and identity covariance matrix. Compute the average ℓ_2 distance of each sample to the origin. Repeat this experiment for $d \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$. What happens to the average distance as d increases? For k-means clustering, what is one implication of this outcome?

(e) [10 MARKS] Consider two d-dimensional hypercubes centered at the origin. The first has a side length of 1 and the second has a side length of 1 - ϵ ($0 < \epsilon < 1$). Give an expression for the ratio of the volume of the second hypercube to the volume of the first (in terms of d and ϵ). What happens as d gets large? How does this help explain the result about average distances from the previous question?