Fall 2020 CS4641/CS7641 A Homework 1

Dr. Mahdi Roozbahani

Deadline: Sep 10, Thursday, 11:59 pm AOE

- No extension of the deadline is allowed. **Late submission will lead to 0 credit**.
- Discussion is encouraged on Piazza as part of the Q/A. However, all assignments **should be done individually**.

Instructions

- This assignment has no programming, only written questions.
- We will be using Gradescope this semester for submission and grading of assignments.
- Your write up must be submitted in PDF form, you may use either Latex or markdown, whichever you prefer. We will not accept handwritten work.
- Please make sure to **start answering each question on a new page**. It makes it more organized to map your answers on GradeScope. When submitting your assignment, you must **correctly map pages of your PDF to each question/subquestion** to reflect where they appear. Improperly mapped questions may not be graded correctly.
- Please **show the calculation process** used to arrive at the answer. Submissions with only the final answer and no derivation/calculation process will receive **0 credit**

1 Linear Algebra [30pts]

1.1 Determinant and Inverse of Matrix [15pts]

Given a matrix M:

$$M = \begin{bmatrix} r & 6 & 0 \\ 2 & 3 & r \\ 4 & 7 & 3 \end{bmatrix}$$

(a) Calculate the determinant of M in terms of r. [4pts]

(b) For what value(s) of r does M^{-1} not exist? Why? What does it mean in terms of rank and singularity of M for these values of r? [3pts]

(c) Calculate M^{-1} by hand for $r = 4$. [5pts] (**Hint 1:** Please double check your answer and make sure $MM^{-1} = I$)

(d) Find the determinant of M^{-1} for $r = 4$. [3pts]

1.2 Characteristic Equation [5pts]

Consider the eigenvalue problem:

$$Ax = \lambda x, x \neq 0$$

where x is a non-zero eigenvector and λ is eigenvalue of A. Prove that the determinant $|A - \lambda I| = 0$.

1.3 Eigenvalues and Eigenvectors [10pts]

Given a matrix A:

$$A = \begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix}$$

(a) Calculate the eigenvalues of A as a function of x [5 pts]

(b) Find the normalized eigenvectors of matrix A [5 pts]

2 Expectation, Co-variance and Independence [18pts]

Suppose X, Y and Z are three different random variables. Let X obey a Bernouli Distribution. The probability disbribution function is

$$p(x) = \begin{cases} 0.5 & x = c \\ 0.5 & x = -c. \end{cases}$$

c is a constant here. Let Y obey a standard Normal (Gaussian) distribution, which can be written as $Y \sim N(0, 1)$. X and Y are independent. Meanwhile, let $Z = XY$.

(a) Show that Z also follows a Normal (Gaussian) distribution. Calculate the Expectation and Variance of Z. [9pts] (**Hint:** Sum rule and conditional probability formula)

(b) How should we choose c such that Y and Z are uncorrelated(which means $Cov(Y, Z) = 0$)? [5pts]

(c) Are Y and Z independent? Make use of probabilities to show your conclusion. Example: $P(Y \in (-1, 0))$ and $P(Z \in (2c, 3c))$ [4pts]

3 Optimization [15 pts]

Optimization problems are related to minimizing a function (usually termed loss, cost or error function) or maximizing a function (such as the likelihood) with respect to some variable x. The Kuhn-Tucker conditions are first-order conditions that provide a unified treatment of constraint optimization. In this question, you will be solving the following optimization problem:

$$\begin{aligned} \max_{x,y} \quad & f(x, y) = 2x^2 + 3xy \\ \text{s.t.} \quad & g_1(x, y) = \frac{1}{2}x^2 + y \le 4 \\ & g_2(x, y) = -y \le -2 \end{aligned}$$

(a) Specify the Legrange function [2 pts]

(b) List the KKT conditions [2 pts]

(c) Solve for 4 possibilities formed by each constraint being active or inactive [5 pts]

(d) List all candidate points [4 pts]

(e) Check for maximality and sufficiency [2 pts]

4 Maximum Likelihood [10 + 25 pts]

4.1 Discrete Example [10 pts]

Suppose we have two types of coins, A and B. The probability of a Type A coin showing heads is θ. The probability of a Type B coin showing heads is 2θ. Here, we have a bunch of coins of either type A or B. Each time we choose one coin and flip it. We do this experiment 10 times and the results are shown in the chart below. (**Hint:** The probabilities aforementioned are for the particular sequence below.)

Coin Type	Result
A	Tail
A	Tail
A	Tail
A	Tail
A	Tail
A	Head
A	Head
B	Head
B	Head
B	Head

(a) What is the likelihood of the result given θ? [4pts]

(b) What is the maximum likelihood estimation for θ? [6pts]

4.2 Normal distribution [15 pts](Bonus for Undergrads)

Suppose that we observe samples of a known function $g(t) = t^3$ with unknown amplitude θ at (known) arbitrary locations $t_1, \ldots, t_N$, and these samples are corrupted by Gaussian noise. That is, we observe the sequence of random variables

$$X_n = \theta t_n^3 + Z_n, \quad n = 1, \ldots, N$$

where the Z_n are independent and $Z_n \sim \text{Normal}\left(0, \sigma^2\right)$

(a) Given $X_1 = x_1, \ldots, X_N = x_N$, compute the log likelihood function

$$\ell\left(\theta; x_1, \ldots, x_N\right) = \log f_{X_1,\ldots,X_N}\left(x_1, \ldots, x_N; \theta\right) = \log\left(f_{X_1}\left(x_1; \theta\right) f_{X_2}\left(x_2; \theta\right) \cdots f_{X_N}\left(x_N; \theta\right)\right)$$

Note that the X_n are independent (as the last equality is suggesting) but not identically distributed (they have different means). [9pts]

(b) Compute the MLE for θ. [6pts]

4.3 Bonus for undergrads [10 pts]

The C.D.F of independent random variables $X_1, X_2, ..., X_n$ is

$$P(X_i \leq x|\alpha, \beta) = \begin{cases} 0, & x < 0 \\ (\dfrac{x}{\beta})^{\alpha}, & 0 \leq x \leq \beta \\ 1, & x > \beta \end{cases}$$

where $\alpha \geq 0$, $\beta \geq 0$.

(a) Write down the P.D.F of above independent random variables. [4pts]

(b) Find the MLEs of α and β. [6pts]

5 Information Theory [32pts]

5.1 Marginal Distribution [6pts]

Suppose the joint probability distribution of two binary random variables X and Y are given as follows.

$X \| Y$	1	2
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

(a) Show the marginal distribution of X and Y, respectively. [3pts]

(b) Find mutual information for the joint probability distribution in the previous question [3pts]

5.2 Mutual Information and Entropy [19pts]

Given a dataset as below.

$Sr.No.$	Age	$Immunity$	$Travelled?$	$UnderlyingConditions$	$Self-quarantine?$
1	$young$	$high$	no	yes	no
2	$young$	$high$	no	no	no
3	$middleaged$	$high$	no	yes	yes
4	$senior$	$medium$	no	yes	yes
5	$senior$	low	yes	yes	yes
6	$senior$	low	yes	no	no
7	$middleaged$	low	yes	no	yes
8	$young$	$medium$	no	yes	no
9	$young$	low	yes	yes	no
10	$senior$	$medium$	yes	yes	yes
11	$young$	$medium$	yes	no	yes
12	$middleaged$	$medium$	no	no	yes
13	$middleaged$	$high$	yes	yes	yes
14	$senior$	$medium$	no	no	no

We want to decide whether an individual working in an essential services industry should be allowed to work or self-quarantine. Each input has four features (x_1, x_2, x_3, x_4): Age, Immunity, Travelled, Underlying Conditions. The decision (quarantine vs not) is represented as Y.

(a) Find entropy $H(Y)$. [3pts]

(b) Find conditional entropy $H(Y|x_1)$, $H(Y|x_4)$, respectively. [8pts]

(c) Find mutual information $I(x_1, Y)$ and $I(x_4, Y)$ and determine which one (x_1 or x_4) is more informative. [4pts]

(d) Find joint entropy $H(Y, x_3)$. [4pts]

5.3 Entropy Proofs [7pts]

(a) Suppose X and Y are independent. Show that $H(X|Y) = H(X)$. [2pts]

(b) Suppose X and Y are independent. Show that $H(X, Y) = H(X) + H(Y)$. [2pts]

(c) Prove that the mutual information is symmetric, i.e., $I(X, Y) = I(Y, X)$ and $x_i \in X, y_i \in Y$ [3pts]

6 Bonus for All [10 pts]

(a) If a random variable X has a Poisson distribution with mean 8, then calculate the expectation $\mathrm{E}[(X + 2)^2]$ [2 pts]

(b) A person decides to toss a fair coin repeatedly until he gets a head. He will make at most 3 tosses. Let the random variable Y denote the number of heads. Find the variance of Y. [4 pts]

(c) Two random variables X and Y are distributed according to

$$f_{x,y}(x, y) = \begin{cases} (x + y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & otherwise \end{cases}$$

What is the probability P(X+Y $\leq$ 1)? [4 pts]