

Fall 2022 CS4641/CS7641 Homework 1

Dr. Mahdi Roozbahani

Deadline: Friday, September 23rd, 11:59 pm AOE

- No unapproved extension of the deadline is allowed. Late submission will lead to 0 credit.
- Discussion is encouraged on Ed as part of the Q/A. However, all assignments should be done individually.
- Plagiarism is a **serious offense**. You are responsible for completing your own work. You are not allowed to copy and paste, or paraphrase, or submit materials created or published by others, as if you created the materials. All materials submitted must be your own.
- All incidents of suspected dishonesty, plagiarism, or violations of the Georgia Tech Honor Code will be subject to the institute's Academic Integrity procedures. If we observe any (even small) similarities/plagiarisms detected by Gradescope or our TAs, **WE WILL DIRECTLY REPORT ALL CASES TO OSI**, which may, unfortunately, lead to a very harsh outcome. **Consequences can be severe, e.g., academic probation or dismissal, grade penalties, a 0 grade for assignments concerned, and prohibition from withdrawing from the class.**

Instructions

- This assignment has no programming, only written questions.
- We will be using Gradescope for submission and grading of assignments.
- Unless a question explicitly states that no work is required to be shown, you must provide an explanation, justification, or calculation for your answer.
- Your write-up must be submitted in PDF form, you may use either Latex, markdown, or any word processing software. **We will NOT accept handwritten work**. Make sure that your work is formatted correctly, for example submit $\sum_{i=0} x_i$ instead of $\text{sum}_{\{i=0\}} x_i$.
- **A useful video tutorial on LaTeX has been created by our TA team** and can be found [here](#) and an Overleaf document with the commands can be found [here](#).
- Please answer 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. Make sure to map the whole solution for each question/subquestion and NOT just the first page. **Improperly mapped questions may not be graded correctly or may receive point deductions.**
- All assignments should be done individually, each student must write up and submit their own answers.
- **Graduate Students:** You are required to complete any sections marked as Bonus for Undergrads

Point Distribution

Q1: Linear Algebra [43pts]

- 1.1 Determinant and Inverse of a Matrix [15pts]
- 1.2 Characteristic Equation [8pts]
- 1.3 Eigenvalues and Eigenvectors [20pts]

Q2: Covariance, Correlation, and Independence [9pts]

- 2.1 Covariance [5pts]
- 2.2 Correlation [4pts]

Q3: Optimization [19pts: 15pts + 4pts Bonus for All]

Q4: Maximum Likelihood [25pts: 10pts + 15pts Bonus for Undergrads]

- 4.1 Discrete Example [10pts]
- 4.2 Weibull Distribution [15pts Bonus for Undergrads]

Q5: Information Theory [35pts]

- 5.1 Marginal Distribution [6pts]
- 5.2 Mutual Information and Entropy [19pts]
- 5.3 Entropy Proofs [10pts]

Q6: Bonus for All [15pts]

1 Linear Algebra [15pts + 8pts + 20pts]

1.1 Determinant and Inverse of Matrix [15pts]

Given a matrix M :

$$M = \begin{bmatrix} 4 & 2 & 1 \\ -3 & r & 2 \\ 0 & 7 & 1 \end{bmatrix}$$

- (a) Calculate the determinant of M in terms of r . (Calculation process is required) [4pts]
- (b) For what value(s) of r does M^{-1} not exist? Why? What does it mean in terms of rank and singularity for these values of r ? [3pts]
- (c) Will all values of r found in part b allow for a column (row) to be expressed as a linear combination of the other columns (rows) respectively? If yes, provide the linear combination of C_3 for column or the linear combination of R_2 for row; if no, explain why. [3pts]
- (d) Write down M^{-1} for $r = 0$. (Calculation process is **NOT** required.) [2pts]
- (e) Find the determinant of M^{-1} for $r = 0$. What is the relationship between the determinant of M and the determinant of M^{-1} ? [3pts]

1.2 Characteristic Equation [8pts]

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 $|A - \lambda I| = 0$.

Note: There are many ways to solve this problem. You are allowed to use linear algebra properties as part of your solution.

1.3 Eigenvalues and Eigenvectors [5+5+10pts]

1.3.1 Eigenvalues [5pts]

Given a matrix A :

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

- (a) Find an expression for the eigenvalues (λ) of A and solve for λ in the terms given. [4pts]
- (b) Find a simple expression for the eigenvalues if $c = a$. [1pt]

1.3.2 Trace and Eigenvectors [5pts]

A symmetric matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ can be decomposed as

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T = \sum_{n=1}^N \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

Where \mathbf{V} is a matrix whose columns are the eigenvectors of \mathbf{A} , \mathbf{v}_n are the columns of \mathbf{V} and $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of \mathbf{A} . The eigenvectors are orthonormal to each other, i.e., $\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$.

- (a) Show that $\text{trace}(\mathbf{A}) = \sum_{n=1}^N \lambda_n$ [3pts]

NOTE: $\mathbf{v}_i^T \mathbf{v}_j \neq \mathbf{v}_i \mathbf{v}_j^T$

- (b) What is the result of the multiplication $\mathbf{V}^T \mathbf{V}$? Show your work or present an argument. [2pts]

1.3.3 Eigenvalue and Eigenvector Calculations [10pts]

Given a matrix

$$\mathbf{A} = \begin{bmatrix} x & 5 \\ 5 & x \end{bmatrix}$$

- (a) Calculate the eigenvalues of \mathbf{A} as a function of x . (Calculation process required). [3pts]
- (b) Find the normalized eigenvectors of matrix \mathbf{A} (Calculation process required). [7pts]

2 Expectation, Co-variance and Independence [5pts + 4pts]

2.1 Covariance [5pts]

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

$$p(x) = \begin{cases} 0.6 & x = c \\ 0.4 & x = -c \end{cases}$$

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

Calculate the covariance of Y and Z ($Cov(Y, Z)$). Do values of c affect the covariance between Y and Z ? [5pts]

2.2 Correlation Coefficient [4pts]

Let X and Y be independent random variables with $\text{var}(X) = 5$ and $\text{var}(Y) = 15$. We do not know $E[X]$ or $E[Y]$. Let $Z = 3X + 2Y$. What is the correlation coefficient $\rho(X, Z) = \frac{\text{cov}(X, Z)}{\sqrt{\text{var}(X)\text{var}(Z)}}$? If applicable, please round your answer to 3 decimal places. [4pts]

3 Optimization [15pts + 4pts Bonus for All]

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 Karush-Kuhn-Tucker (KKT) conditions are first-order conditions for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. In this question, you will be solving the following optimization problem:

$$\begin{aligned} \max_{x,y} \quad & f(x,y) = -4y + xy \\ \text{s.t.} \quad & g_1(x,y) = 2x^2 + y^2 \leq 12 \\ & g_2(x,y) = x \leq 1 \end{aligned}$$

- (a) Write the Lagrange function for the maximization problem. Now change the maximum function to a minimum function (i.e. $\min_{x,y} f(x,y) = -4y + xy$) and provide the Lagrange function for the minimization problem with the same constraints g_1 and g_2 . [2pts]

Note: The minimization problem is only for part (a).

- (b) List the names of all of the KKT conditions and its corresponding mathematical equations or inequalities for this specific maximization problem [2pts]
- (c) Solve for 4 possibilities formed by each constraint being active or inactive. Do not forget to check the inactive constraints for each point. Candidate points must satisfy the inactive constraints. [5pts]
- (d) List the candidate point(s) (there may be 0, 1, 2, or any number of candidate points) [4pts]
- (e) Find the **one** candidate point for which $f(x,y)$ is largest. Check if $L(x,y)$ is concave or convex at this point by using the [Hessian](#) in the [second partial derivative test](#). [2pts]

HINT 1: Click [here](#) for an example maximization problem.

HINT 2: Click [here](#) to determine how to set up the problem for minimization in part (a).

- (f) **BONUS FOR ALL:** Make a contour plot of objective function $f(x,y)$ and constraints g_1 and g_2 using the template [Google Colab](#) code. Mark the maximum candidate point and include a screenshot of your plot. Also include the text output from the last cell in the Google Colab for grading purposes. Lastly, briefly explain why your plot makes sense in one sentence. [4pts]

Note 1: Points on a line in the contour plot have equal values of the objective function. Keeping this in mind, you should be able to figure out the approximate location of the maximum.

Note 2: To use the Google Colab notebook, click "Copy to Drive" upon initial opening

4 Maximum Likelihood [10pts + 15pts Bonus for Undergrads]

4.1 Discrete Example [10pts]

Marion and Shreeya are arguing over which course they should take in Fall 2022. Marion's argument is that they should take CS-7650 NLP because Professor Roozbahani will teach it. Shreeya's argument is that they should take CS-7641 ML because it would be difficult to take NLP without having introductory knowledge of Machine Learning.

To resolve this conflict, their other friend Nicole makes a proposition that they should leave it to chance to decide which course they should take. Marion then proposes that Shreeya will toss a 6-sided die 6 times, and Shreeya must get anything except 3 during the first 5 times and must get 3 during the 6th time. Any other combination will make Marion the winner. But Shreeya is also allowed to tamper with the die in any manner she likes to increase her odds.

Now, Shreeya needs you to help her have her way. If the probability of getting a 3 is θ and the probability of landing on 1 is double of that of landing on 2, 4, 5, and 6, what value of θ is most likely to ensure that they will have to take CS-7641 ML? Use your expertise of Maximum Likelihood Estimation and probability distribution function to convince Shreeya.

NOTE: You must specify the log-likelihood function and use MLE to solve this problem for full credit. You may assume that the log-likelihood function is concave for this question

4.2 Weibull distribution [15pts Bonus for Undergrads]

The Weibull distribution is defined as

$$P(X = x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0$$

- (a) Assume we have one observed data x_1 , and $X_1 \sim Weibull(\lambda)$, what is the likelihood given λ and k ? [2 pts]
- (b) Now, assume we are given n such values (x_1, \dots, x_n) , $(X_1, \dots, X_n) \sim Weibull(\lambda)$. Here X_1, \dots, X_n are i.i.d. random variables. What is the likelihood of this data given λ and k ? You may leave your answer in product form. [3 pts]
- (c) What is the maximum likelihood estimator of λ ? [10 pts]

5 Information Theory [6pts + 19pts + 10pts]

5.1 Marginal Distribution [6pts]

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

X \ Y	0	1
0	$\frac{1}{16}$	$\frac{1}{4}$
1	$\frac{3}{16}$	$\frac{1}{2}$

- (a) Show the marginal distribution of X and Y , respectively. [3pts]
- (b) Find mutual information $I(X, Y)$ for the joint probability distribution in the previous question to at least 3 decimal places (please use base 2 to compute logarithm) [3pts]

5.2 Mutual Information and Entropy [19pts]

A recent study has shown symptomatic infections are responsible for higher transmission rates. Using the [data](#) collected from positively tested patients, we wish to determine which feature(s) have the greatest impact on whether or not some will present with symptoms. To do this, we will compute the entropies, conditional entropies, and mutual information of select features. Please use base 2 when computing logarithms.

ID	Age Group (x_1)	Vaccine Doses (x_2)	Wears Mask? (x_3)	Underlying Conditions (x_4)	Symptomatic (Y)
1	Y	H	F	T	T
2	Y	H	F	F	F
3	A	H	F	T	T
4	S	M	F	T	T
5	S	L	T	T	T
6	S	L	T	F	F
7	A	L	T	F	T
8	Y	L	F	T	F
9	Y	L	T	T	F
10	S	M	T	T	T

Table 1: Age Groups: {(Y)outh, (A)dult, (S)enior}, Vaccine Doses: {(H) booster, (M) 2 doses, (L) 1 dose}

- Find entropy $H(Y)$ to at least 3 decimal places. [3pts]
- Find conditional entropy $H(Y|x_2)$, $H(Y|x_4)$, respectively, to at least 3 decimal places. [8pts]
- Find mutual information $I(x_2, Y)$ and $I(x_4, Y)$ and determine which one (x_2 or x_4) is more informative. [4pts]
- Find joint entropy $H(Y, x_3)$ to at least 3 decimal places. [4pts]

5.3 Entropy Proofs [10pts]

- (a) Write the discrete case mathematical definition for $H(X|Y)$ and $H(X)$. [3pts]
- (b) **Using the mathematical definition of $H(X)$ and $H(X|Y)$ from part (a)**, prove that $I(X; Y) = 0$ if X and Y are independent. (Note: you must provide a mathematical proof and cannot use the visualization shown in class [found here](#))

Start from $I(X; Y) = H(X) - H(X|Y)$ [7pts]

6 Bonus for All [15 pts]

- (a) X, Y are two independent $N(0, 1)$ random variables, and we have random variables P, Q defined as

$$P = 3X + XY^2$$

$$Q = X$$

then calculate the variance $Var(P + Q)$ [5pts]

- (b) Suppose that X and Y have joint pdf given by

$$f_{X,Y}(x, y) = \begin{cases} 2e^{-2y}, & 0 \leq x \leq 1, y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

What are the marginal probability density functions for X and Y ? [5 pts]

- (c) A person decides to toss a biased coin with $P(heads) = 0.2$ repeatedly until he gets a head. He will make at most 5 tosses. Let the random variable Y denote the number of heads. Find the variance of Y . [5 pts]