CS4641 - Summer 2022 Assignment 1

Shihui Liu

TOTAL POINTS

139 / 142

QUESTION 1

Q1 Linear Algebra 43 pts

1.1 1.1 Determinant and Inverse of a Matrix 15 / 15

✓ - **0 pts** Correct

- **15 pts** No submission
- **1 pts** Incorrectly/missing assigned pages

Error in part a

- **1 pts** Incorrect determinant
- **3 pts** No work shown
- **4 pts** No attempt or missing solution

Error in part b

- **0.5 pts** Did not discuss rank
- **0.5 pts** Did not discuss singularity
- **1 pts** One value of r is wrong
- **2 pts** Both values of r are wrong
- **3 pts** No attempt or missing solution

Error in part c - Only 1 linear combination required

- **2 pts** Incorrect answer on r & linear combinations
- **1 pts** Incorrect linear combination
- **1 pts** Example of linear combination not provided, but argued through another method (e.g. RREF has zeros in last row, similar rows after row operations etc.)
- **2 pts** Did not provide any linear combination
- **3 pts** No attempt or missing solution

Error in part d

- **1 pts** Incorrect inverse, but work shown
- **2 pts** Incorrect inverse, no work shown
- **2 pts** No attempt or missing solution

Error in part e

- **1.5 pts** Incorrect determinant for $$M^{-1}$$
- **1.5 pts** Incorrect relationship or did not discuss the relationship between determinant of $$M$$ and $$M^{-1}$$
- **3 pts** No attempt or missing solution
- **0.5 pts** Incorrect Equation

1.2 1.2 Characteristic Equation 8 / 8

✓ - **0 pts** Correct

- **1 pts** Incorrectly/missing assigned pages
- **2 pts** Generally correct but missing one or two steps
- **3 pts** Does not sufficiently explain why $$|A-\lambda I| = 0$$
- **4 pts** Did not explain why $$|A-\lambda I| = 0$$
- **4 pts** Vague or circular definition or proof
- **8 pts** No attempt or missing solution
- **1 pts** Minor error

1.3 1.3.1 Eigenvalues 5 / 5

✓ - **0 pts** Correct

- **5 pts** Missing solution
- **1 pts** Incorrectly/missing assigned pages

Error in part a

- **1 pts** Minor error(s) in calculation/final answer
- **2 pts** Major error(s) in calculation/final answer
- **2 pts** Did not solve for $$\lambda$$
- **2.5 pts** No work shown
- **4 pts** No attempt or missing solution

Error in part b

- **0.25 pts** one $$\lambda$$ expression is incorrect
- **0.5 pts** Incorrect or missing $$\lambda$$ expression
- **1 pts** No attempt or missing solution

- **0.5 pts** Didn't solve for $$\lambda$$

1.4 1.3.2 Eigenvectors 5 / 5

✓ - **0 pts** Correct

- **5 pts** Missing solution
- **1 pts** Incorrectly/missing assigned pages

Error in part a

- **1 pts** One or two minor errors/missing steps
- **2 pts** Several incorrect or missing steps
- **2.5 pts** Circular logic or incorrect solution
- **3 pts** No attempt or missing solution

Error in part b

- **0.5 pts** Incorrect or missing answer for $$V^TV$$
- **0.5 pts** Minor error or missing step
- **1 pts** Several errors or missing steps
- **1.5 pts** Circular or incorrect argument
- **2 pts** No work or argument presented
- **2 pts** No attempt or missing solution

1.5 1.3.3 Eigenvalue and Eigenvector Calculations (BONUS FOR ALL) 10 / 10

✓ - **0 pts** Correct

- **10 pts** Missing solution
- **1 pts** Incorrectly/missing assigned pages

Error in part a

- **1 pts** Incorrect eigenvalues
- **3 pts** No work or explanation shown
- **3 pts** No attempt or missing solution

Error in part b

- **1 pts** Did not normalize eigenvectors at all or correctly
- **2 pts** Small error in calculating eigenvectors
- **4 pts** Incomplete work for calculating eigenvectors
- **5 pts** Incorrect eigenvectors and approach
- **7 pts** No work shown
- **7 pts** No attempt or missing solution

QUESTION 2

Q2 Covariance, Correlation, and Independence 9 pts

2.1 2.1 Covariance 5 / 5

✓ - **0 pts** Correct

- **5 pts** No attempt or missing solution
- **1 pts** Minor error
- **1 pts** Incorrect answer about values of c
- **2 pts** Missing answer about values of $$c$$
- **2 pts** Incorrect value for $$Cov(Y,Z)$$
- **3 pts** Missing calculation for $$Cov(Y,Z)$$
- **2 pts** Insufficient work shown
- **1 pts** Incorrect pages

2.2 2.2 Correlation 3 / 4

- **0 pts** Correct
- **1 pts** Incorrectly/missing assigned pages

✓ - **1 pts** Incorrect or missing value of $$\rho(X,Z)$$

- **1 pts** Minor calculation error
- **2 pts** Major calculation error
- **3 pts** No work shown
- **4 pts** Missing solution

QUESTION 3

3 Q3 Optimization 13 / 15

- **0 pts** Correct
- **15 pts** Missing solution
- **1 pts** Incorrectly/missing assigned pages
- **1 pts** Wrong answer for Q3.a
- **1 pts** Q3.b: lack KKT conditions
- **2 pts** Wrong answer for Q3.c
- **1 pts** Wrong answer for Q3.d
- **1 pts** Wrong answer for Q3.e

Error in part a

- **1 pts** Incorrect Lagrange function
- **2 pts** No attempt or missing solution
- **1 pts** Need more computational process
- **4 pts** Wrong answer for Q3.d

Error in part b (correctness based on answer in part a)

- **0 pts** Correct KKT conditions based on Lagrange function in part a

- **0.5 pts** Missing or incorrect stationarity condition
- **0.5 pts** Missing or incorrect complementary slackness condition
- **0.5 pts** Missing or incorrect primal feasibility condition
- **0.5 pts** Missing or incorrect dual feasibility condition
- **2 pts** No attempt or missing solution

Error in part c

- **1 pts** Error in $$\lambda_1$$ active and $$\lambda_2$$ active
- **1 pts** Error in $$\lambda_1$$ active and $$\lambda_2$$ inactive
- **1 pts** Error in $$\lambda_1$$ inactive and $$\lambda_2$$ active
- **1 pts** Error in $$\lambda_1$$ inactive and $$\lambda_2$$ inactive
- **2.5 pts** Work shown but does not explore all 4 possibilities
- **2.5 pts** Generally correct method but incorrect calculations due to initial error in Lagrange function
- **4 pts** Errors in method and incorrect calculations due to initial error in Lagrange function
- **4 pts** No work shown solving for 4 possibilities
- **5 pts** No attempt or missing solution

Error in part d (correctness based on answer in part c)

- **0 pts** Correct candidate points based on part c calculations with incorrect initial Lagrange function
- **1 pts** One extra incorrect candidate point
- **2 pts** Missing correct candidate point
- **2 pts** Two or more extra incorrect candidate points
- **4 pts** No attempt or missing solution

Error in part e

- **0 pts** Correct minimizing point based on listed points in part d
- **1 pts** Incorrect minimizing point
- **1 pts** Missing calculations for Hessian and Second Partial Derivative Test
- **1 pts** Missing or incorrect convexity for $$L(x,y)$$
- **2 pts** No attempt or missing solution

- **2 pts** No answer for Q3.b
- **5 pts** No answer for Q3.c
- **4 pts** No answer for Q3.d
- ✓ **- 2 pts** No answer for Q3.e
- **1 pts** See comments for more details

QUESTION 4

Q4 Maximum Likelihood 25 pts

4.1 4.1 Discrete Example 10 / 10

- ✓ **- 0 pts** Correct
- **1 pts** Incorrectly/missing assigned pages
- **1 pts** Minor error
- **2 pts** Work shown, incorrect final answer.
- **3 pts** Inadequate work shown. Final answer correct.
- **4 pts** Inadequate work shown. Final answer incorrect.
- **8 pts** No work shown. Correct Answer.
- **10 pts** No work shown. Incorrect Answer.
- **10 pts** Missing submission.
- **3 pts** Wrong or missed log-likelihood function.
- **3 pts** Wrong or missed derivative of the log-likelihood function.
- **2 pts** Did not set the derivative to zero
- **9 pts** No work shown, approach step is correct.
- **2 pts** Incorrect derivative function

4.2 4.2 Poisson Distribution (BONUS FOR UG) 15 / 15

- ✓ **- 0 pts** Correct
- **15 pts** Not attempted or did not map the corresponding page. (Submit a regrade if you are the latter case.)
- **1 pts** Incorrectly/missing assigned pages

4.2 a

- **0 pts** Correct
- **1 pts** work shown, incorrect likelihood.
- **2 pts** No work, incorrect likelihood.
- **2 pts** Not attempted.

- **1 pts** Expression missing

4.2 b

- **0 pts** Correct
- **0.5 pts** minor error
- **1 pts** Work shown, incorrect likelihood.
- **2 pts** No work, incorrect likelihood.
- **3 pts** Not attempted.

4.2 c

- **0 pts** Click here to replace this description.
- **1 pts** minor error
- **2 pts** work shown, incorrect MLE.
- **3 pts** Inadequate work shown. Correct MLE.
- **5 pts** Inadequate work shown. Incorrect MLE.
- **8 pts** No work. Correct answer.
- **10 pts** Missing submission.
- **8 pts** Missing steps and incorrect approach

QUESTION 5

Q5 Information Theory 35 pts

5.1 5.1 Marginal Distribution 6 / 6

✓ **- 0 pts** Correct

- **1 pts** Incorrectly/missing assigned pages
- **6 pts** Not attempted/ No work shown.

Error in part a

- **0.5 pts** (a) Minor Error
- **1.5 pts** (a) Work shown, incorrect answer.
- **1.5 pts** Incorrect marginal distribution for X OR Y
- **3 pts** (a) Left Empty/Incorrect with no work shown

Error in part b

- **0.5 pts** Minor math error
- **1 pts** (b) Partial work showed, Correct Answer
- **1.5 pts** (b) Work shown, but incorrect/missing final answer
- **3 pts** (b) Left Empty/Incorrect with no work shown

5.2 5.2 Mutual Information and Entropy 19 / 19

✓ **- 0 pts** Correct

- **1 pts** Incorrectly/missing assigned pages
- **5 pts** Used log10 instead of log2
- **19 pts** No submission

error in part a

- **2 pts** 5.2.a Work shown but Incorrect Final Answer
- **3 pts** 5.2.a Incorrect/Missing H(Y)

error in part b

- **1 pts** Mostly correct method with minor error
- **2 pts** Missing step
- **4 pts** Incorrect answer, work shown
- **4 pts** 5.2.b One incorrect conditional entropy
- **8 pts** 5.2.b Incorrect/Missing conditional entropies for both x's

error in part c

- **1 pts** Math error
- **1 pts** 5.2.c Incorrect/missing claim of which x is more informative
- **1.5 pts** 5.2.c One incorrect mutual information value
- **2 pts** 5.2.c Incorrect mutual informations, but work shown
- **3 pts** Missing mutual information but correct claim
- **4 pts** 5.2.c Both incorrect mutual informations and incorrect/missing decision on which x is more informative

error in part d

- **1 pts** Minor Error
- **2 pts** Incorrect answer, work shown
- **2 pts** 5.2.d Incorrect mutual entropy but correct formula setup
- **4 pts** 5.2.d Incorrect/Missing Mutual Entropy

5.3 5.3 Entropy Proofs 10 / 10

✓ **- 0 pts** Correct

- **1 pts** Incorrectly/missing assigned pages

error in part a

- **1 pts** (a) Error in definition or not rigorous/formal enough
- **2 pts** (a) Error in definitions
- **3 pts** (a) Incorrect definitions/ Not attempted.

error in part b

- **- 1 pts** (b) Minor error or missing step in proof.
- **- 2 pts** (b) Errors or missing steps in proof
- **- 4 pts** (b) Error in proof. Refer to comment.
- **- 7 pts** (b) Incorrect/ Not attempted.

QUESTION 6

6 Q6 BONUS FOR ALL **15 / 15**

- ✓ **- 0 pts Correct**
- **- 1 pts** Incorrectly/missing assigned pages
- **- 15 pts** Not attempted.
- **- 2.5 pts** (a) Partially correct.
- **- 2.5 pts** (b) Partially correct.
- **- 2.5 pts** (c) Partially correct.
- **- 5 pts** (a) Incorrect
- **- 5 pts** (b) Incorrect
- **- 5 pts** (c) Incorrect

1 Linear Algebra [15pts + 8pts + 10pts + 10pts Bonus for All]

1.1 Determinant and Inverse of Matrix [15pts]

Given a matrix M:

$$M = \begin{bmatrix} 5 & r & 2 \\ 1 & -2 & -3 \\ 5 & 0 & r \end{bmatrix}$$

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

$$\begin{aligned} det \begin{bmatrix} 5 & r & 2 \\ 1 & -2 & -3 \\ 5 & 0 & r \end{bmatrix} &= 5 \cdot \begin{bmatrix} -2 & -3 \\ 0 & r \end{bmatrix} - r \cdot \begin{bmatrix} 1 & -3 \\ 5 & r \end{bmatrix} + 2 \cdot \begin{bmatrix} 1 & -2 \\ 5 & 0 \end{bmatrix} \\ &= 5 \cdot (-2r - 0) - r \cdot (r - (-15)) + 2 \cdot (0 - (-2) \cdot 5) \\ &= -10r - r^2 - 15r + 20 \\ &= -r^2 - 25r + 20 \end{aligned}$$

(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]

We know that by definition of square matrices, a square matrix A is invertible if and only if the determinant of the square matrix $A \neq 0$. We can see from the calculation below that there are two possible values in which the determinant of $M = 0$

$$\begin{aligned} 0 &= -r^2 - 25r + 20 \\ r_{1,2} &= \frac{-(-25) \pm \sqrt{(-25)^2 - 4(-1) \cdot 20}}{2 \cdot (-1)} \\ r_{1,2} &= -\frac{25 \pm \sqrt{705}}{2} \end{aligned}$$

$$r_1 = -\frac{25 + \sqrt{705}}{2} \qquad r_2 = -\frac{25 - \sqrt{705}}{2}$$

(c) Will all values of r found in part b allow for a row (or a column) to be expressed as a linear combination of the other rows (or columns) respectively? If yes, provide just one linear combination example, if no, explain why. [3pts]

$$M = \begin{bmatrix} 5 & -\frac{25+\sqrt{705}}{2} & 2 \\ 1 & -2 & -3 \\ 5 & 0 & -\frac{25+\sqrt{705}}{2} \end{bmatrix},$$

By row reducing the above matrix, we get:

$$M = \begin{bmatrix} 1 & 0 & -\frac{25+\sqrt{705}}{10} \\ 0 & 1 & -\frac{\sqrt{705}-5}{10} \\ 0 & 0 & 0 \end{bmatrix},$$

We can express column 3 as a linearly combination of column 1 and 2:

$$x_3 = (-\frac{25 + \sqrt{705}}{10})x_1 + (-\frac{\sqrt{705} - 5}{10})x_2$$

(d) Write down M^{-1} for $r = 0$. (Calculation process is **NOT** required.) [2pts]

$$M^-1 = \begin{bmatrix} 0 & 0 & \frac{1}{5} \\ -\frac{3}{4} & -\frac{1}{2} & \frac{17}{20} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}$$

(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]

$$\begin{aligned} det \begin{bmatrix} 0 & 0 & \frac{1}{5} \\ -\frac{3}{4} & -\frac{1}{2} & \frac{17}{20} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix} &= 0 \cdot \begin{bmatrix} -\frac{1}{2} & \frac{17}{20} \\ 0 & -\frac{1}{2} \end{bmatrix} + 0 \cdot \begin{bmatrix} -\frac{3}{4} & \frac{17}{20} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} + \frac{1}{5} \cdot \begin{bmatrix} -\frac{3}{4} & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \\ &= 0 \cdot \frac{1}{4} - 0 \cdot (\frac{3}{8} - \frac{17}{40}) + \frac{1}{5} \cdot \frac{1}{4} \\ &= \frac{1}{20} \end{aligned}$$

Set $r = 0$, the determinant of M is equal to 20. Therefore, the relationship is that $|M^{-1}| = \frac{1}{|M|}$

1.1 1.1 Determinant and Inverse of a Matrix 15 / 15

✓ **- 0 pts** **Correct**

- 15 pts No submission

- 1 pts Incorrectly/missing assigned pages

Error in part a

- 1 pts Incorrect determinant

- 3 pts No work shown

- 4 pts No attempt or missing solution

Error in part b

- 0.5 pts Did not discuss rank

- 0.5 pts Did not discuss singularity

- 1 pts One value of r is wrong

- 2 pts Both values of r are wrong

- 3 pts No attempt or missing solution

Error in part c - Only 1 linear combination required

- 2 pts Incorrect answer on r & linear combinations

- 1 pts Incorrect linear combination

- 1 pts Example of linear combination not provided, but argued through another method (e.g. RREF has zeros in last row, similar rows after row operations etc.)

- 2 pts Did not provide any linear combination

- 3 pts No attempt or missing solution

Error in part d

- 1 pts Incorrect inverse, but work shown

- 2 pts Incorrect inverse, no work shown

- 2 pts No attempt or missing solution

Error in part e

- 1.5 pts Incorrect determinant for $$M^{-1}$$

- 1.5 pts Incorrect relationship or did not discuss the relationship between determinant of $$M$$ and $$M^{-1}$$

- 3 pts No attempt or missing solution

- 0.5 pts Incorrect Equation

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.

First, we can move λx to the left hand side of the equation:

$Ax - \lambda x = 0$

Since we know that a vector multiplied by an identity matrix is itself: $x = x\lambda$, we can rewrite the equation above to:

$Ax - \lambda Ix = 0$

Using the distributive property of matrix multiplication, we can write the following:

$(A - \lambda I)x = 0$

Since by definition, eigenvector x cannot be $0(A0 = 0 = \lambda 0$, which produces undefined eigenvalue). Furthermore, $(A - \lambda I)x = 0$ has a non-zero solution if and only if $(A - \lambda I)$ has a non-empty null space. This could only happen if $(A - \lambda I)$ is singular, which means that:

$|A - \lambda I| = 0$

1.2 1.2 Characteristic Equation 8 / 8

- ✓ **- 0 pts** Correct
- **- 1 pts** Incorrectly/missing assigned pages
- **- 2 pts** Generally correct but missing one or two steps
- **- 3 pts** Does not sufficiently explain why $$|A-\lambda I| = 0$$
- **- 4 pts** Did not explain why $$|A-\lambda I| = 0$$
- **- 4 pts** Vague or circular definition or proof
- **- 8 pts** No attempt or missing solution
- **- 1 pts** Minor error

1.3 Eigenvalues and Eigenvectors [5pts + 5pts + 10pts Bonus for All]

1.3.1 [5pts]

Given a matrix A:

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

(a) Find an expression for the eigenvalues (λ) of $\boldsymbol{A}$ and solve for λ in the terms given. [4pts]

We use the equation $|A - \lambda I| = 0$:

$$\begin{aligned} det \begin{bmatrix} a-\lambda & b \\ b & c-\lambda \end{bmatrix} &= 0 \\ (a-\lambda)(c-\lambda) - b^2 &= 0 \\ ac - a\lambda - c\lambda + \lambda^2 - b^2 &= 0 \\ \lambda^2 - (a+c)\lambda + ac - b^2 &= 0 \end{aligned}$$

Solve the equation above:

$$\lambda_{1,2} = \frac{a+c \pm \sqrt{(a+c)^2 - 4 \cdot 1 \cdot (ac - b^2)}}{2 \cdot 1}$$

$$\lambda_{1,2} = \frac{a+c \pm \sqrt{a^2 + 2ac + c^2 - 4ac + 4b^2}}{2}$$

$$\lambda_{1,2} = \frac{a+c \pm \sqrt{a^2 - 2ac + c^2 + 4b^2}}{2}$$

$$\lambda_{1,2} = \frac{a+c \pm \sqrt{a^2 - 2ac + c^2 + 4b^2}}{2}$$

$$\lambda_1 = \frac{a+c+\sqrt{a^2 - 2ac + c^2 + 4b^2}}{2} \qquad \lambda_2 = \frac{a+c-\sqrt{a^2 - 2ac + c^2 + 4b^2}}{2}$$

(b) Find a simple expression for the eigenvalues if $c = a$. [1pt]

$$\lambda_{1,2} = \frac{a+a \pm \sqrt{a^2 - 2a \cdot a + a^2 + 4b^2}}{2}$$

$$\lambda_{1,2} = \frac{2a \pm \sqrt{4b^2}}{2}$$

$$\lambda_{1,2} = \frac{2a \pm 2b}{2}$$

$$\lambda_{1,2} = a \pm b$$

$$\lambda_1 = a + b \qquad \lambda_2 = a - b$$

1.3 1.3.1 Eigenvalues 5 / 5

✓ **- 0 pts** **Correct**

- 5 pts Missing solution

- 1 pts Incorrectly/missing assigned pages

Error in part a

- 1 pts Minor error(s) in calculation/final answer

- 2 pts Major error(s) in calculation/final answer

- 2 pts Did not solve for $$\lambda$$

- 2.5 pts No work shown

- 4 pts No attempt or missing solution

Error in part b

- 0.25 pts one $$\lambda$$ expression is incorrect

- 0.5 pts Incorrect or missing $$\lambda$$ expression

- 1 pts No attempt or missing solution

- 0.5 pts Didn't solve for $$\lambda$$

1.3.2 Eigenvalues [5pts]

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

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

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

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

From the Eigen Decomposition Theorem, we know that V is a square matrix since A can be written as $V\Lambda V^T$. Also, trace has the property that if ABC is a sqaure matrix, $tr(ABC) = tr(BCA) = tr(CAB)$. Since V, Λ and V^T are all square matrix:

$$trace(A) = trace(V\Lambda V^T) = trace(VV^T\Lambda)$$

From the square matrix property, we know that $UU^T = I$. Therefore, we can rewrite the equation above to:

$$trace(VV^T\Lambda) = trace(I\Lambda) = trace(\Lambda)$$

Since Λ is a diagonal matrix whose elements are the eigenvalues of A, its trace is simply the sum of all the eigenvalues of A: $trace(\Lambda) = \sum_{n=1}^{N} \lambda_n$ Therefore, we have shown that:

$$trace(A) = trace(V\Lambda V^T) = trace(VV^T\Lambda) = trace(\Lambda) = \sum_{n=1}^{N} \lambda_n$$

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

We know that V is a matrix whose columns are the eigenvectors of A. We can write V as $\left[v_1, v_2, v_3, ..., v_n\right]$and V^T as $V^T = \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \\ ... \\ v_n^T \end{bmatrix}$. We can then write VV^T as:

$$VV^T = \left[v_1, v_2, v_3, ..., v_n\right] \begin{bmatrix} v_1^T \\ v_2^T \\ v_3^T \\ ... \\ v_n^T \end{bmatrix} = \begin{bmatrix} v_1v_1^T & v_2v_1^T & ... & v_nv_1^T \\ v_1v_2^T & v_2v_2^T & ... & v_nv_2^T \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ v_1v_n^T & v_2v_2^T & ... & v_nv_n^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{bmatrix} = I$$

The result of multiplication of V^TV is the identity matrix I

1.4 1.3.2 Eigenvectors 5 / 5

✓ **- 0 pts** **Correct**

- 5 pts Missing solution

- 1 pts Incorrectly/missing assigned pages

Error in part a

- 1 pts One or two minor errors/missing steps

- 2 pts Several incorrect or missing steps

- 2.5 pts Circular logic or incorrect solution

- 3 pts No attempt or missing solution

Error in part b

- 0.5 pts Incorrect or missing answer for $$V^TV$$

- 0.5 pts Minor error or missing step

- 1 pts Several errors or missing steps

- 1.5 pts Circular or incorrect argument

- 2 pts No work or argument presented

- 2 pts No attempt or missing solution

1.3.3 Eigenvectors [10pts - Bonus for All]

Given a matrix

$$\boldsymbol{A} = \begin{bmatrix} x & 4 \\ 4 & x \end{bmatrix}$$

(a) Calculate the eigenvalues of $\boldsymbol{A}$ as a function of x. (Calculation process required). [3pts]

$$\begin{aligned}(A - \lambda I) &= \begin{bmatrix} x-\lambda & 4 \\ 4 & x-\lambda \end{bmatrix} \\ &= (x-\lambda)^2 - 4 \cdot 4 \\ &= x^2 - 2\lambda x + \lambda^2 - 16 \\ &= \lambda^2 - 2\lambda x + x^2 - 16 \\ &= (-\lambda + x + 4)(-\lambda + x - 4)\end{aligned}$$

$\lambda_1 = x + 4$ $\qquad\qquad$ $\lambda_2 = x - 4$

(b) Find the normalized eigenvectors of matrix $\boldsymbol{A}$ (Calculation process required). [7pts]

$$\begin{aligned}\lambda_1 &= x + 4 \\ (A - \lambda_1 I) &= \begin{bmatrix} x-\lambda_1 & 4 \\ 4 & x-\lambda_1 \end{bmatrix} \\ &= \begin{bmatrix} x-x-4 & 4 \\ 4 & x-x-4 \end{bmatrix} \\ &= \begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 \\ 4 & -4 \end{bmatrix} (R1 = -R1/4) \\ &= \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} (R2 = R2 - 4R1)\end{aligned}$$

$$\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Assume $x_1 = t$, then $x_2 = t$,

$$\begin{bmatrix} t-t \\ 0-0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\therefore x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} t, v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Eigenvector $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, the normalized eigenvector is:

$$\begin{aligned}\frac{v_1}{|v_1|} &= \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{1^2 + 1^2}} \\ &= \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{2}} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\end{aligned}$$

$$
\begin{aligned}
\lambda_2 &= x - 4 \\
(A - \lambda_2 I) &= \begin{bmatrix} x - \lambda_2 & 4 \\ 4 & x - \lambda_2 \end{bmatrix} \\
&= \begin{bmatrix} x - x + 4 & 4 \\ 4 & x - x + 4 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 \\ 4 & 4 \end{bmatrix} (R1 = R1/4) \\
&= \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} (R2 = R2 - 4R1) \\
\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\end{aligned}
$$

Assume $x_1 = -t$, then $x_2 = t$,

$$
\begin{aligned}
\begin{bmatrix} -t + t \\ 0 - 0 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\therefore x &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} t, v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}
\end{aligned}
$$

Eigenvector $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, the normalized eigenvector is:

$$
\begin{aligned}
\frac{v_1}{|v_1|} &= \frac{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}{\sqrt{(-1)^2 + 1^2}} \\
&= \frac{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}{\sqrt{2}} \\
&= \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}
\end{aligned}
$$

1.5 1.3.3 Eigenvalue and Eigenvector Calculations (BONUS FOR ALL) 10 / 10

✓ **- 0 pts** **Correct**

- 10 pts Missing solution

- 1 pts Incorrectly/missing assigned pages

Error in part a

- 1 pts Incorrect eigenvalues

- 3 pts No work or explanation shown

- 3 pts No attempt or missing solution

Error in part b

- 1 pts Did not normalize eigenvectors at all or correctly

- 2 pts Small error in calculating eigenvectors

- 4 pts Incomplete work for calculating eigenvectors

- 5 pts Incorrect eigenvectors and approach

- 7 pts No work shown

- 7 pts No attempt or missing solution

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

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

c is a constant here. 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 ($Cov(Y, Z)$) and determine whether values of c would affect the correlation. [5pts]

Follow the definition of covariance, we can write $Cov(Y, Z)$ as $Cov(Y, Z) = E(YZ) - E(Y)E(Z)$. Since we know the Y is a Gaussian Distribution with $\mu = 0$, $E(Y) = \mu = 0$. Therefore we can write the covariance equation to: $Cov(Y, Z) = E(YZ) - 0 \cdot E(Z) = E(YZ)$.

Since we know that $Z = XY$, we can further rewrite the covariance equation to: $Cov(YZ) = Cov(XY^2)$. We can calculate further as follows:

$$\begin{aligned} Cov(Y, Z) &= E[XY^2] \\ &= P(X = c)E[XY^2|X = c] + P[X = -c]E[XY^2|X = -c] \\ &= 0.5E[cY^2] + 0.5E[-cY^2] \text{ (By definition of X)} \\ &= 0.5cE[Y^2] - 0.5cE[Y^2] \text{ (By linearity of Expectation)} \\ &= 0 \end{aligned}$$

As we see from the calculation above, $Cov(Y, Z) = 0$, and Y and Z are uncorrolated with each other. We also see from the last step of the calculation that the value of c does not affect the correlation, since the covariance is always 0 regardless of the value of c.

2.1 2.1 Covariance **5 / 5**

- ✓ **- 0 pts** **Correct**
- **- 5 pts** No attempt or missing solution
- **- 1 pts** Minor error
- **- 1 pts** Incorrect answer about values of c
- **- 2 pts** Missing answer about values of $$c$$
- **- 2 pts** Incorrect value for $$Cov(Y,Z)$$
- **- 3 pts** Missing calculation for $$Cov(Y,Z)$$
- **- 2 pts** Insufficient work shown
- **- 1 pts** Incorrect pages

2.2 Correlation Coefficient [4pts]

Let X and Y be independent random variables with $var(X) = 4$ and $var(Y) = 12$. We do not know $E[X]$ or $E[Y]$. Let $Z = 3X + Y$. What is the correlation coefficient $\rho(X, Z) = \frac{cov(X,Z)}{\sqrt{var(X)var(Z)}}$? [4pts]

First, we consider the covariance $Cov(X, Z)$:

$$\begin{aligned}
Cov(X, Z) &= Cov(X, 3X + Y) \text{ (Given Z=3X+Y)} \\
&= Cov(X, 3X) + Cov(X, Y) \text{ (Distributive Property of Covariance)} \\
&= Cov(X, 3X) + 0 \text{ (X and Y are independent)} \\
&= Cov(X, 3X) \\
&= 3Cov(X, X) \text{ (Take out the constant)} \\
&= 3Var(X) \text{ (The covariance of a random variable with itself is its variance)} \\
&= 3 \cdot 4 \text{ (Given Var(X) = 4)} \\
&= 12
\end{aligned}$$

Then, we consider $var(Z)$:

$$\begin{aligned}
Var(Z) &= Var(3X + Y) \text{ (Given Z=3X+Y)} \\
&= Var[3X] + Var[Y] + 2 \cdot Cov[X, Y] \\
&= Var[3X] + Var[Y] + 0 \text{ (X and Y are independent)} \\
&= 3Var[X] + Var[Y] \text{ (Property of Variance, take out constant)} \\
&= 3 \cdot 4 + 12 \text{ (Given Var(X) = 4, Var(Y) = 12))} \\
&= 24
\end{aligned}$$

We have shown that $Cov(X, Z) = 12$, $Var(X) = 4$ and $Var(Z) = 24$. We can then calculate the correlation coefficient:

$$\begin{aligned}
\rho(X, Z) &= \frac{cov(X, Z)}{\sqrt{var(X)var(Z)}} \\
&= \frac{12}{\sqrt{4 \cdot 24}} \\
&= \frac{12}{\sqrt{96}} \\
&= \frac{12}{4\sqrt{6}} \\
&= \frac{3}{\sqrt{6}}
\end{aligned}$$

2.2 2.2 Correlation 3 / 4

- **- 0 pts** Correct
- **- 1 pts** Incorrectly/missing assigned pages
- ✓ **- 1 pts Incorrect or missing value of $$\rho(X,Z)$$**
- **- 1 pts** Minor calculation error
- **- 2 pts** Major calculation error
- **- 3 pts** No work shown
- **- 4 pts** Missing solution

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 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:

$$\begin{aligned} \min_{x,y} \ & f(x,y) = 7x^2 + 4y \\ \text{s.t.} \ & g_1(x,y) = x^2 + y^2 \leq 2 \\ & g_2(x,y) = x \leq 1 \end{aligned}$$

(a) Write the specific Lagrange function for this minimization problem [2pts]

The Lagrange Function can be written as:

$$\begin{aligned} L(x,y,\lambda_1,\lambda_2) &= f(x,y) + \lambda_1 g_1(x,y) + \lambda_2 g_2(x,y) \\ &= (7x^2 + 4y) + \lambda_1(x^2 + y^2 - 2) + \lambda_2(x - 1) \end{aligned}$$

(b) List the names of all of the KKT conditions and its corresponding mathematical equations or inequalities for this specific minimization problem [2pts]

1. Stationary Condition

$$\begin{aligned} \frac{\partial L(x,y,\lambda_1,\lambda_2)}{\partial x} &= 14x + 2\lambda_1 x + \lambda_2 = 0 \\ \frac{\partial L(x,y,\lambda_1,\lambda_2)}{\partial y} &= 4 + 2\lambda_1 y = 0 \end{aligned}$$

2. Primal feasibility

$$\begin{aligned} x^2 + y^2 - 2 \leq 0 \\ x - 1 \leq 0 \end{aligned}$$

3. Dual feasibility

$$\begin{aligned} \lambda_1 \geq 0 \\ \lambda_2 \geq 0 \end{aligned}$$

4. Complementary slackness

$$\begin{aligned} \lambda_1(x^2 + y^2 - 2) = 0 \\ \lambda_2(x - 1) = 0 \end{aligned}$$

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

Case 1: $x = 1, \lambda_2 > 0$.
Check for stationary condition, we have $14 + 2\lambda_1 + \lambda_2 = 0$ and $4 + 2\lambda_1 y = 0$. This cannot be satisfied since from the first term we can see that for λ_2 to be positive, λ_1 must be negative, and vise versa.

Case 2: $\lambda_1 = 0$(Dual feasibility inactive)
Check for stationary condition, we have $14x + \lambda_2 = 0$ and $4 = 0$, which is not possible.

Case 3: $\lambda_2 = 0$(Dual feasibility inactive)
Check for stationary condition, $14x + 2\lambda_1 x = 0$ and $4 + 2\lambda_1 y = 0$. A satisfactory solution would be $x = 0, y = -\sqrt{2}, \lambda_1 = \sqrt{2}, \lambda_2 = 0$.

Case 4: $x^2 + y^2 = 2$
In this case, we can parameterize x and y in terms of θ and consider $x = \sqrt{2}cos(\theta)$ and $y = \sqrt{2}sin(\theta)$. λ_1 and λ_2 are positive within a certain rage of values of x and y, therefore satisfies the stationary condition.

(d) List the candidate point(s) (there may be 0, 1, 2, or any number of candidate points) [4pts]
$(x = 0, y = -\sqrt{2})$
$(x = \sqrt{2}cos(\theta), y = \sqrt{2}sin(\theta))$

(e) Find the **one** candidate point for which f(x,y) is smallest. 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 a maximization example.
HINT 2: Click here to determine how to set up the problem for minimization in parts (a) and (b).

3 Q3 Optimization **13 / 15**

- **- 0 pts** Correct
- **- 15 pts** Missing solution
- **- 1 pts** Incorrectly/missing assigned pages
- **- 1 pts** Wrong answer for Q3.a
- **- 1 pts** Q3.b: lack KKT conditions
- **- 2 pts** Wrong answer for Q3.c
- **- 1 pts** Wrong answer for Q3.d
- **- 1 pts** Wrong answer for Q3.e

Error in part a

- **- 1 pts** Incorrect Lagrange function
- **- 2 pts** No attempt or missing solution
- **- 1 pts** Need more computational process
- **- 4 pts** Wrong answer for Q3.d

Error in part b (correctness based on answer in part a)

- **- 0 pts** Correct KKT conditions based on Lagrange function in part a
- **- 0.5 pts** Missing or incorrect stationarity condition
- **- 0.5 pts** Missing or incorrect complementary slackness condition
- **- 0.5 pts** Missing or incorrect primal feasibility condition
- **- 0.5 pts** Missing or incorrect dual feasibility condition
- **- 2 pts** No attempt or missing solution

Error in part c

- **- 1 pts** Error in $$\lambda_1$$ active and $$\lambda_2$$ active
- **- 1 pts** Error in $$\lambda_1$$ active and $$\lambda_2$$ inactive
- **- 1 pts** Error in $$\lambda_1$$ inactive and $$\lambda_2$$ active
- **- 1 pts** Error in $$\lambda_1$$ inactive and $$\lambda_2$$ inactive
- **- 2.5 pts** Work shown but does not explore all 4 possibilities
- **- 2.5 pts** Generally correct method but incorrect calculations due to initial error in Lagrange function
- **- 4 pts** Errors in method and incorrect calculations due to initial error in Lagrange function
- **- 4 pts** No work shown solving for 4 possibilities
- **- 5 pts** No attempt or missing solution

Error in part d (correctness based on answer in part c)

- **- 0 pts** Correct candidate points based on part c calculations with incorrect initial Lagrange function
- **- 1 pts** One extra incorrect candidate point
- **- 2 pts** Missing correct candidate point
- **- 2 pts** Two or more extra incorrect candidate points
- **- 4 pts** No attempt or missing solution

Error in part e

- **- 0 pts** Correct minimizing point based on listed points in part d

- **- 1 pts** Incorrect minimizing point
- **- 1 pts** Missing calculations for Hessian and Second Partial Derivative Test
- **- 1 pts** Missing or incorrect convexity for $$L(x,y)$$
- **- 2 pts** No attempt or missing solution

- **- 2 pts** No answer for Q3.b
- **- 5 pts** No answer for Q3.c
- **- 4 pts** No answer for Q3.d
- ✓ **- 2 pts** **No answer for Q3.e**
- **- 1 pts** See comments for more details

4 Maximum Likelihood [10pts + 15pts]

4.1 Discrete Example [10pts]

Devshree and Angana are arguing over where they should go for winter break. Devshree's argument is that they should go to Boston because Megha, their childhood friend, lives there and thus they will get to have a good time. Angana's argument is that they should go to Florida because Florida is very near to Atlanta, it has the mildest winter and on the other hand, Boston's weather is very rough during the winter.

To resolve this conflict, their other friend Yusuf makes a proposition that they should leave it to chance to decide where they should spend their winter break. Devshree then proposes that Angana will toss a 6-sided die 5 times, and Angana must get anything except 2 during the first 4 times and must get 2 during the 5th time. Any other combination will make Devshree the winner. But Angana is also allowed to tamper with the die in any manner she likes to increase her odds.

Now, Angana needs you to help her have her way. If the probability of getting a 2 is θ and the probabilities of landing on any other number is the same, what value of θ is most likely to ensure that they will have to go to Florida? Use your expertise of Maximum Likelihood Estimation and probability distribution function to convince Angana.

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

$$
\begin{aligned}
L(p) &= \frac{5}{6}(1-\theta) \times \frac{5}{6}(1-\theta) \times \frac{5}{6}(1-\theta) \times \frac{5}{6}(1-\theta) \times \frac{1}{6}(\theta) \\
&= \frac{5^4}{6^5}(1-\theta)^4\theta \\
l(\theta) &= log(\frac{5^4}{6^5}(1-\theta)^4\theta) \\
&= log(5^4) - log(6^5) + log((1-\theta)^4) + log(\theta)) \\
&= log(5^4) - log(6^5) + 4log((1-\theta)) + log(\theta)) \\
\frac{dl(\theta)}{d\theta} &= -\frac{4}{1-\theta} + \frac{1}{\theta} = 0 \\
0 &= -4\theta + 1 - \theta \\
5\theta &= 1 \\
\theta &= \frac{1}{5}
\end{aligned}
$$

The maximum likelihood for θ is $\frac{1}{5}$

4.1 4.1 Discrete Example 10 / 10

✓ **- 0 pts** **Correct**

- 1 pts Incorrectly/missing assigned pages

- 1 pts Minor error

- 2 pts Work shown, incorrect final answer.

- 3 pts Inadequate work shown. Final answer correct.

- 4 pts Inadequate work shown. Final answer incorrect.

- 8 pts No work shown. Correct Answer.

- 10 pts No work shown. Incorrect Answer.

- 10 pts Missing submission.

- 3 pts Wrong or missed log-likelihood function.

- 3 pts Wrong or missed derivative of the log-likelihood function.

- 2 pts Did not set the derivative to zero

- 9 pts No work shown, approach step is correct.

- 2 pts Incorrect derivative function

4.2 Poisson distribution [15 pts]: Bonus for undergrads

The Poisson distribution is defined as

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}(k = 0, 1, 2, ...).$$

(a) Assume we have one observed data x_1, and $X_1 \sim Poisson(\lambda)$, what is the likelihood given λ? [2 pts]

$$L(\lambda; x_1) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}$$

(b) Now, assume we are given n such values $(x_1, ..., x_n)$, $(X_1, ..., X_n) \sim Poisson(\lambda)$. Here $X_1, ..., X_n$ are i.i.d. random variables. What is the likelihood of this data given λ? You may leave your answer in product form. [3 pts]

$$L(\lambda; x_1) = \prod_{j=1}^{n} \frac{e^{-\lambda}\lambda^{x_j}}{x_j!}$$

(c) What is the maximum likelihood estimator of λ? [10 pts]

Calculate natural log likelihood function:

$$\begin{aligned}
l(\lambda; x_1...x_n) &= ln(\prod_{j=1}^{n} \frac{e^{-\lambda}\lambda^{x_j}}{x_j!}) \\
&= \sum_{j=1}^{n} ln(\frac{e^{-\lambda}\lambda^{x_j}}{x_j!}) \text{ (Log product rule)} \\
&= \sum_{j=1}^{n} [ln(\lambda^{x_j}) + ln(e^{-\lambda}) - ln(x_j!)] \text{ (Log quotient rule)} \\
&= \sum_{j=1}^{n} [x_j ln(\lambda) - \lambda - ln(x_j!)] \text{ (Log power rule)} \\
&= -n\lambda + ln(\lambda)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} ln(x_j!) \text{ (Take out constant)}
\end{aligned}$$

Calculate derivative with respect to λ:

$$\begin{aligned}
\frac{d}{d\lambda}(-n\lambda + ln(\lambda)\sum_{j=1}^{n} x_j - \sum_{j=1}^{n} ln(x_j!)) &= -n + \frac{1}{\lambda}\sum_{j=1}^{n} x_j - 0 \\
&= -n + \frac{1}{\lambda}\sum_{j=1}^{n} x_j
\end{aligned}$$

Solve for λ to get the maximum likelihood estimator:

$$0 = -n + \frac{1}{\lambda}\sum_{j=1}^{n} x_j$$

$$n = \frac{1}{\lambda}\sum_{j=1}^{n} x_j$$

$$\lambda = \frac{1}{n}\sum_{j=1}^{n} x_j$$

4.2 4.2 Poisson Distribution (BONUS FOR UG) **15 / 15**

✓ **- 0 pts** **Correct**

- 15 pts Not attempted or did not map the corresponding page. (Submit a regrade if you are the latter case.)

- 1 pts Incorrectly/missing assigned pages

4.2 a

- 0 pts Correct

- 1 pts work shown, incorrect likelihood.

- 2 pts No work, incorrect likelihood.

- 2 pts Not attempted.

- 1 pts Expression missing

4.2 b

- 0 pts Correct

- 0.5 pts minor error

- 1 pts Work shown, incorrect likelihood.

- 2 pts No work, incorrect likelihood.

- 3 pts Not attempted.

4.2 c

- 0 pts Click here to replace this description.

- 1 pts minor error

- 2 pts work shown, incorrect MLE.

- 3 pts Inadequate work shown. Correct MLE.

- 5 pts Inadequate work shown. Incorrect MLE.

- 8 pts No work. Correct answer.

- 10 pts Missing submission.

- 8 pts Missing steps and incorrect approach

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	$\frac{1}{8}$	$\frac{3}{8}$
1	0	$\frac{1}{2}$

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

For X, the marginal distribution for $x = 0$ is $P_X(x = 0) = (\frac{1}{8} + \frac{3}{8}) = \frac{4}{8} = \frac{1}{2}$. The marginal distribution for $x = 1$ is $P_X(x = 1) = 0 + \frac{1}{2} = \frac{1}{2}$.
For Y, the marginal distribution for $y = 0$ is $P_Y(y = 0) = \frac{1}{8} + 0 = \frac{1}{8}$. The marginal distribution for $y = 1$ is $P_Y(y = 1) = \frac{3}{8} + \frac{1}{2} = \frac{7}{8}$.

X \ Y	0	1	marginal
0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$
1	0	$\frac{1}{2}$	$\frac{1}{2}$
marginal	$\frac{1}{8}$	$\frac{7}{8}$	

(b) Find mutual information for the joint probability distribution in the previous question to at least 3 decimal places (please use base 2 to compute the logarithm) [3pts]

According to the definition of mutual information given in class:

$$
\begin{aligned}
I(X,Y) &= \sum_{x,y} P(x,y) \cdot log_2 \frac{P(x,y)}{P(x)P(y)} \\
&= P(x=0,y=0) \cdot log_2 \frac{P(x=0,y=0)}{P(x=0)P(y=0)} + P(x=0,y=1) \cdot log_2 \frac{P(x=0,y=1)}{P(x=0)P(y=1)} \\
&+ P(x=1,y=0) \cdot log_2 \frac{P(x=1,y=0)}{P(x=1)P(y=0)} + P(x=1,y=1) \cdot log_2 \frac{P(x=1,y=1)}{P(x=1)P(y=1)} \\
&= \frac{1}{8} \cdot log_2 \frac{\frac{1}{8}}{\frac{1}{2} \cdot \frac{1}{8}} + \frac{3}{8} \cdot log_2 \frac{\frac{3}{8}}{\frac{1}{2} \cdot \frac{7}{8}} + 0 \cdot log_2 \frac{0}{\frac{1}{2} \cdot \frac{1}{8}} + \frac{1}{2} \cdot log_2 \frac{\frac{1}{2}}{\frac{1}{2} \cdot \frac{7}{8}} \\
&= 0.138
\end{aligned}
$$

5.1 5.1 Marginal Distribution 6 / 6

✓ **- 0 pts** Correct

- **- 1 pts** Incorrectly/missing assigned pages
- **- 6 pts** Not attempted/ No work shown.

Error in part a

- **- 0.5 pts** (a) Minor Error
- **- 1.5 pts** (a) Work shown, incorrect answer.
- **- 1.5 pts** Incorrect marginal distribution for X OR Y
- **- 3 pts** (a) Left Empty/Incorrect with no work shown

Error in part b

- **- 0.5 pts** Minor math error
- **- 1 pts** (b) Partial work showed, Correct Answer
- **- 1.5 pts** (b) Work shown, but incorrect/missing final answer
- **- 3 pts** (b) Left Empty/Incorrect with no work shown

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*	*yes*
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 does not quarantine) is represented as Y. Please use base 2 when you compute the logarithm.

(a) Find entropy $H(Y)$ to at least 3 decimal places. [3pts]

We see from the self-quarantine column that the number of **yes** is 9, and number of **no** is 5. Therefore, $p(y = yes) = \frac{9}{14}$, $p(y = no) = \frac{5}{14}$

$$
\begin{aligned}
H(Y) &= -\sum_i p_i log_2 p_i \\
&= -\frac{9}{14} log_2(\frac{9}{14}) - \frac{5}{14} log_2(\frac{5}{14}) \\
&= 0.940285958671
\end{aligned}
$$

(b) Find conditional entropy $H(Y|x_2)$, $H(Y|x_4)$, respectively, to at least 3 decimal places. [8pts]

We know from the dataset $P(X_2 = high) = \frac{4}{14}$,$P(X_2 = medium) = \frac{6}{14}$ and $P(X_2 = high) = \frac{4}{14}$. We can also construct the following probability table for $P(Y|x_2)$:

Self-quarantine \ Immunity	high	medium	low
yes	$\frac{3}{4}$	$\frac{4}{6}$	$\frac{2}{4}$
no	$\frac{1}{4}$	$\frac{2}{6}$	$\frac{2}{4}$

$$
\begin{aligned}
H(Y|X_2) &= \sum_x P(X_2 = x_2) H(Y|X_2 = x_2) \\
&= P(X_2 = high) H(Y|X_2 = high) + P(X_2 = medium) H(Y|X_2 = medium) \\
&+ P(X_2 = high) H(Y|X_2 = medium) \\
&= \frac{4}{14} H(\frac{3}{4}, \frac{1}{4}) + \frac{6}{14} H(\frac{4}{6}, \frac{2}{6}) + \frac{4}{14} H(\frac{2}{4}, \frac{2}{4}) \\
&= \frac{4}{14} \cdot (\frac{3}{4} \log_2(\frac{1}{\frac{3}{4}}) + \frac{1}{4} \log_2(\frac{1}{\frac{1}{4}})) + \frac{6}{14} \cdot (\frac{4}{6} \log_2(\frac{1}{\frac{4}{6}}) + \frac{2}{6} \log_2(\frac{1}{\frac{2}{6}})) + \frac{4}{14} \cdot (\frac{2}{4} \log_2(\frac{1}{\frac{2}{4}}) + \frac{2}{4} \log_2(\frac{1}{\frac{2}{4}})) \\
&= 0.911063393012
\end{aligned}
$$

We know from the dataset $P(X_4 = yes) = \frac{8}{14}$ and $P(X_2 = no) = \frac{6}{14}$. We can also construct the following probability table for $P(Y|x_4)$:

Self-quarantine \ Underlying Cond	yes	no
yes	$\frac{6}{8}$	$\frac{3}{6}$
no	$\frac{2}{8}$	$\frac{3}{6}$

$$
\begin{aligned}
H(Y|X_4) &= \sum_x P(X_4 = x_4)H(Y|X_4 = x_4) \\
&= P(X_4 = yes)H(Y|X_4 = yes) + P(X_4 = no)H(Y|X_4 = no) \\
&= \frac{8}{14}H(\frac{6}{8}, \frac{2}{8}) + \frac{6}{14}H(\frac{3}{6}, \frac{3}{6}) \\
&= \frac{8}{14} \cdot (\frac{6}{8}log_2(\frac{1}{\frac{6}{8}}) + \frac{2}{8}log_2(\frac{1}{\frac{2}{8}})) + \frac{6}{14} \cdot (\frac{3}{6}log_2(\frac{1}{\frac{3}{6}}) + \frac{3}{6}log_2(\frac{1}{\frac{3}{6}})) \\
&= 0.892158928262
\end{aligned}
$$

(c) 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]

We know that the definition of mutual information is:

$$I(X_i, Y) = H(Y) - H(Y|X_i)$$

From part a and b, we know the values of $H(Y)$ and $H(Y|x_2)$. Therefore, we can calculate $I(x_2, Y)$:

$$
\begin{aligned}
I(x_2, Y) &= H(Y) - H(Y|x_2) \\
&= 0.940285958671 - 0.911063393012 \\
&= 0.029222565659
\end{aligned}
$$

From part a and b, we know the values of $H(Y)$ and $H(Y|x_4)$. Therefore, we can calculate $I(x_4, Y)$:

$$
\begin{aligned}
I(x_4, Y) &= H(Y) - H(Y|x_4) \\
&= 0.940285958671 - 0.892158928262 \\
&= 0.0481270304083
\end{aligned}
$$

From the calculation above, we can see that $I(x_2, Y) < I(x_4, Y)$. The more the reduction in entropy, the more informative a feature. Since x_4 has more reduction , x_4 is more informative.

(d) Find joint entropy $H(Y, x_3)$ to at least 3 decimal places. [4pts]
We first construct the following probability table from the given dateset.

Self-quarantine \ Travelled	yes	no
yes	$\frac{5}{14}$	$\frac{4}{14}$
no	$\frac{2}{14}$	$\frac{3}{14}$

$$
\begin{aligned}
H(Y, X_3) &= \sum_{y,x_3} P(Y = y, X_3 = x_3)log\frac{1}{P(Y = y, X_3 = x_3)} \\
&= \frac{5}{14}log_2\frac{1}{\frac{5}{14}} + \frac{4}{14}log_2\frac{1}{\frac{4}{14}} + \frac{2}{14}log_2\frac{1}{\frac{2}{14}} + \frac{3}{14}log_2\frac{1}{\frac{3}{14}} \\
&= 1.9241743523
\end{aligned}
$$

5.2 5.2 Mutual Information and Entropy **19 / 19**

- ✓ **- 0 pts** Correct
- **- 1 pts** Incorrectly/missing assigned pages
- **- 5 pts** Used log10 instead of log2
- **- 19 pts** No submission

error in part a

- **- 2 pts** 5.2.a Work shown but Incorrect Final Answer
- **- 3 pts** 5.2.a Incorrect/Missing H(Y)

error in part b

- **- 1 pts** Mostly correct method with minor error
- **- 2 pts** Missing step
- **- 4 pts** Incorrect answer, work shown
- **- 4 pts** 5.2.b One incorrect conditional entropy
- **- 8 pts** 5.2.b Incorrect/Missing conditional entropies for both x's

error in part c

- **- 1 pts** Math error
- **- 1 pts** 5.2.c Incorrect/missing claim of which x is more informative
- **- 1.5 pts** 5.2.c One incorrect mutual information value
- **- 2 pts** 5.2.c Incorrect mutual informations, but work shown
- **- 3 pts** Missing mutual information but correct claim
- **- 4 pts** 5.2.c Both incorrect mutual informations and incorrect/missing decision on which x is more informative

error in part d

- **- 1 pts** Minor Error
- **- 2 pts** Incorrect answer, work shown
- **- 2 pts** 5.2.d Incorrect mutual entropy but correct formula setup
- **- 4 pts** 5.2.d Incorrect/Missing Mutual Entropy

5.3 Entropy Proofs [10pts]

(a) Write the discrete case mathematical definition for $H(X|Y)$ and $H(X)$. [3pts]

The definition of entropy of a discrete random variable $H(X)$ is defined as:

$$H(X) = \sum_{x} p(x) log_2 \frac{1}{p(x)}$$

The definition of conditional entropy of X given Y is:

$$H(X|Y) = \sum_{y \in Y} p(y_i) log_2 p(X|Y = y_i) = \sum_{x \in X, y \in Y} p(x_i, y_i) log \frac{p(y_i)}{p(y_i, x_i)}$$

(b) **Using the mathematical definition of $H(X)$ and $H(X|Y)$ from part (a)**, prove that $H(X|Y) = H(X)$ if X and Y are independent. (Note: you have to prove it and cannot use the visualization shown in class found here [7pts]

$$\begin{aligned}
H(X|Y) &= \sum_{x \in X, y \in Y} p(x_i, y_i) log \frac{p(y_i)}{p(y_i, x_i)}) \\
&= \sum_{x \in X, y \in Y} p(x_i, y_i) log \frac{p(y_i)}{p(y_i)p(x_i)} \text{(Since X and Y are independent, } p(x, y) = p(x)p(y)) \\
&= \sum_{x \in X, y \in Y} p(x_i, y_i) log \frac{1}{p(x_i)} \\
&= \sum_{x \in X} p(x_i) log \frac{1}{p(x_i)} \text{ (Sum rule)} \\
&= H(X) \text{ (Definition of entropy)}
\end{aligned}$$

5.3 5.3 Entropy Proofs **10 / 10**

✓ **- 0 pts** Correct

- 1 pts Incorrectly/missing assigned pages

error in part a

- 1 pts (a) Error in definition or not rigorous/formal enough

- 2 pts (a) Error in definitions

- 3 pts (a) Incorrect definitions/ Not attempted.

error in part b

- 1 pts (b) Minor error or missing step in proof.

- 2 pts (b) Errors or missing steps in proof

- 4 pts (b) Error in proof. Refer to comment.

- 7 pts (b) Incorrect/ Not attempted.

6 Bonus for All [15 pts]

(a) If a random variable X has a Poisson distribution with mean 5, then calculate the expectation E$[(2X-3)^2]$ [5 pts]

Since the distribution has a mean of 5, we know that $E(x) = 5$. We also know that for a Poisson distribution, mean is the same as variance, therefore: $Var(x) = 5 = E[X^2] - E[X]^2$.We then rewrite the expectation:

$$\begin{aligned}
E[(2X-3)^2] &= E[4X^2 - 12X + 9] \text{ (Expand term)} \\
&= E[4X^2] - E[12X] + E[9] \text{ (Property of Expectation)} \\
&= 4E[X^2] - 12E[X] + 9 \\
&= 4(5 + E[X]^2) - 25E[X] + 9 \text{ (Derived from variance showed above)} \\
&= 20 + 4E[X]^2 - 25E[X] + 9 \\
&= 20 + 4(5^2) - 25 \cdot 5 + \\
&= 69
\end{aligned}$$

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

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

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

Marginal Probability Density Function for $(X|0 \le x \le 1)$:

$$\begin{aligned}
f(x) &= \int_0^\infty 2e^{-2y} dy \\
&= 2\int_0^\infty e^{-2y} dy \\
&= 2\int_0^\infty e^u - \frac{1}{2} du \; (u = -2y, \frac{du}{dy} = -2) \\
&= -\int_0^\infty e^u du \\
&= -\int_0^\infty e^{-2y} dy \\
&= -e^{-2\infty} - (-e^{-2 \cdot 0}) \\
&= 0 + 1 \\
&= 1
\end{aligned}$$

Marginal Probability Density Function for $(X|x < 0 \text{ and } x > 1)$:

$$f(x) = \int 0 dy = 0 \tag{1}$$

Marginal Probability Density Function for $(Y|y \ge 0)$:

$$\begin{aligned}
f(y) &= \int_0^1 2e^{-2y} dx \\
&= 2e^{-2y} \int_0^1 dx \\
&= 2e^{-2y} x|_0^1 \\
&= 2e^{-2y}(1-0) \\
&= 2e^{-2y}
\end{aligned}$$

Marginal Probability Density Function for $(Y|y < 0)$:

$$f(y) = \int 0dx = 0$$

(c) A person decides to toss a biased coin with $P(heads) = 0.4$ 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. [5 pts]

We know from the given that $P(H) = 0.4$, therefore $P(T) = 1 - 0.4 = 0.6$.
Consider the first scenario that he makes one toss and gets a head. The probability for this scenario is: $P(H) = 0.4$.
Consider the second scenario that he makes one toss and gets a tail, and he makes another toss and gets a head. The probability for this scenario is $P(TH) = 0.6 \cdot 0.4 = 0.24$.
Consider the third scenario that he makes three tosses. He gets tails for the first two tosses, and gets a head for the third one. The probability for this scenario is$P(TTH) = 0.6 \cdot 0.6 \cdot 0.4 = 0.144$.
Consider the forth scenario that he makes three tosses and all three are tails. The probability for this scenario is $P(TTT) = 0.6 \cdot 0.6 \cdot 0.6 = 0.216$
We can calculate the variance using the formula $Var(X) = E[X^2] - E[X]^2$.
Expectation:

$$\begin{aligned}
E[X] &= \sum_i x_i p(x_i) \\
&= (1 \cdot 0.4) + (1 \cdot 0.24) + (1 \cdot 0.144) + (0 \cdot 0.216) \\
&= 0.784 \\
E[X^2] &= \sum_i x_i^2 p(x_i) \\
&= (1^2 \cdot 0.4) + (1^2 \cdot 0.24) + (1^2 \cdot 0.144) + (0^2 \cdot 0.216) \\
&= 0.784 \\
Var[X] &= E[X^2] - E[X]^2 \\
&= 0.784 - (0.784)^2 \\
&= 0.169344
\end{aligned}$$

6 Q6 BONUS FOR ALL 15 / 15

✓ **- 0 pts** **Correct**

- 1 pts Incorrectly/missing assigned pages

- 15 pts Not attempted.

- 2.5 pts (a) Partially correct.

- 2.5 pts (b) Partially correct.

- 2.5 pts (c) Partially correct.

- 5 pts (a) Incorrect

- 5 pts (b) Incorrect

- 5 pts (c) Incorrect

Summer 2022 CS4641/CS7641 A Homework 1

Dr. Mahdi Roozbahani

Deadline: Wednesday, June 8th, 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 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: 33pts + 10pts Bonus for All]

- 1.1 Determinant and Inverse of a Matrix [15pts]
- 1.2 Characteristic Equation [8pts]
- 1.3 Eigenvalues and Eigenvectors [20pts: 5pts + 5pts + 10pts Bonus for All]

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

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

Q3: Optimization [15pts]

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

- 4.1 Discrete Example [10pts]
- 4.2 Poisson 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]