ML-4641 Fall22 Assignment 1

Jun Yeop Kim

TOTAL POINTS

90 / 146

QUESTION 1

Q1 Linear Algebra 43 pts

1.1 1.1 Determinant and Inverse of a Matrix 14.5 / 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 or incorrect rank

✓ **- 0.5 pts** *Did not discuss singularity or incorrect singularity*

- **- 2 pts** Value of r is wrong
- **- 2 pts** Value of r is wrong
- **- 3 pts** No attempt or missing solution
- **- 2 pts** Value of r is missing.

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 or missing 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

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 Trace and Eigenvectors **3.5 / 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 **6 / 10**

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

Error in part a

- **1 pts** Incorrect eigenvalue(s)
- **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 or missing one or two steps

✓ **- 4 pts** *Incomplete or incorrect work for calculating eigenvectors*

- **5 pts** Incorrect eigenvectors and approach
- **7 pts** No work shown
- **7 pts** No attempt or missing solution
- **0.5 pts** The normalized eigenvectors for first lambda are actually representing the same thing.

- **0 pts** Click here to replace this description.

QUESTION 2

Q2 Covariance, Correlation, and Independence 9 pts

2.1 2.1 Covariance **2 / 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 Coefficient **4 / 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

Q3 Optimization 19 pts

3.1 Q3 a to e **8 / 15**

- **0 pts** Correct
- **15 pts** Missing solution
- **1 pts** Incorrectly/missing assigned pages

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 or Incorrect answer

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 $$L(x,y)$$

- **1 pts** Incorrect Hessian values

- **1 pts** No comment on convexity

- **2 pts** No attempt or missing solution

3.2 Q3 f (BONUS FOR ALL) 0 / 4

- **0 pts** Correct

✓ **- 4 pts** *Did not attempt/no solution*

- **2 pts** No code/incorrect graph

- **1 pts** No explanation

- **0 pts** Incorrect graph due to issues in previous sections

QUESTION 4

Q4 Maximum Likelihood 25 pts

4.1 4.1 Discrete Example 7 / 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

No log-likelihood function is mentioned. Only likelihood.

4.2 4.2 Weibull Distribution (BONUS FOR UG) 0 / 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.
- **3 pts** Work shown for one term, not the product term
- **5 pts** Inadequate work shown. Incorrect MLE.
- **8 pts** No work. Correct answer.
- **8 pts** Missing steps and incorrect approach
- **9 pts** No work. Incorrect answer.
- **10 pts** Missing submission.

QUESTION 5

Q5 Information Theory 35 pts

5.1 5.1 Marginal Distribution **6 / 6**

✓ **- 0 pts** *Correct*

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

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 **16 / 19**

- **0 pts** Correct
- **19 pts** Not attempted/No work shown
- **1 pts** Incorrectly/missing assigned pages
- **5 pts** Used log10 instead of log2

error in part a

- **1 pts** Minor Error
- **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*

- 10 pts Not attempted/No work shown

- 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 **0 / 15**

- 0 pts Correct

✓ **- 15 pts** *Not attempted.*

- 1 pts Incorrectly/missing assigned pages

- 1 pts a) Almost correct

- 1 pts b) minor incorrect

- 4 pts (a)Minor Correct

- 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 + 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] **Solution:**

$$4 * \begin{vmatrix} r & 2 \\ 7 & 1 \end{vmatrix} - (-3) * \begin{vmatrix} 2 & 1 \\ 7 & 1 \end{vmatrix} + 0 = 4 * (r - 14) + 3 * (-5)$$

$= 4 * r - 71$

(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]

Solution: Value of M^-1 does not exist for $r = 71$. The reason is from (a), we know that $det(A) = 0$ when $r = 71$. The determinant of A being 0 means that the matrix is a linearly dependent matrix, so $rank(A) \neq 3$ and the matrix is not singular.

(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]

Solution: For the column part,

$$a_1 * \begin{bmatrix} 4 \\ -3 \\ 0 \end{bmatrix} + a_2 * \begin{bmatrix} 2 \\ r \\ 7 \end{bmatrix} = a_3 * \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$4a_1 + 2a_2 = a_3$
$-3a_1 + \frac{71}{4}a_2 = 2a_3$
$7a_2 = a_3$

From the equation above, we can derive $a_1 = \frac{5}{4}a_2, 7a_2 = a_3$, which gives us
$5C_1 + 4C_2 = 28C_3$ as we consider a_1, a_2, a_3 as C_1, C_2, C_3

For the row part,

$$a_1 * \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} + a_2 * \begin{bmatrix} 0 \\ 7 \\ 1 \end{bmatrix} = a_3 * \begin{bmatrix} -3 \\ \frac{71}{4} \\ 2 \end{bmatrix}$$

$4a_1 = -3a_2$
$2a_1 + 7a_3 = \frac{71}{4}a_2$
$a_1 + a_3 = a_2$

From the euqation above, we can derive $a_1 = -3, a_2 = 4, a_3 = 11$, which gives us
$-3R_1 + 11R_3 = 4R_2$ as we consider a_1, a_2, a_3 as R_1, R_2, R_3

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

Solution:

$$M^-1 = \frac{1}{71} * \begin{bmatrix} 14 & -5 & -4 \\ -3 & -4 & 11 \\ 21 & 28 & -6 \end{bmatrix}$$

(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]

Solution:

$$det(M^{-}1) = \frac{1}{71^3} * (14 * \begin{vmatrix} -4 & 11 \\ 28 & -6 \end{vmatrix} + 3 * \begin{vmatrix} -5 & -4 \\ 28 & -6 \end{vmatrix} + 21 * \begin{vmatrix} -5 & -4 \\ -4 & 11 \end{vmatrix})$$

$$= \frac{1}{71^3} * (14 * (24 - 28 * 11) + 3 * (30 + 112) + 21 * (-55 - 16)) = \frac{-1}{71}$$

Since r is 0, $det(M) * det(M^{-}1) = 1$, which shows the inversely proportional relationship.

1.1 1.1 Determinant and Inverse of a Matrix **14.5 / 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 or incorrect rank

✓ **- 0.5 pts** *Did not discuss singularity or incorrect singularity*

- **2 pts** Value of r is wrong
- **2 pts** Value of r is wrong
- **3 pts** No attempt or missing solution
- **2 pts** Value of r is missing.

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 or missing 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

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.

Solution: I can proceed the proof using contradiction. I would assume that $det(A - \lambda * I \neq 0$.

By definition, if $det(A - \lambda * I \neq 0$, then it means matrix $|A - \lambda I|$ is linearly independent. To be linearly independent, the only case $(A - \lambda I) * x = 0$ is $x = 0$, since $A - \lambda I$ means A is a full rank matrix. This contradicts the basis that x is $x \neq 0$.
Therefore, the assumption was wrong and the statement that $det(A - \lambda I)$ is 0 is true.

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 [5+5+10pts]

1.3.1 Eigenvalues [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]

Solution: $A - \lambda I = 0$

This can be expressed as

$$(A - \lambda I) = \begin{bmatrix} a - \lambda & b \\ b & c - \lambda \end{bmatrix}$$

$$det(A - \lambda I) = (a - \lambda) * (c - \lambda) - b^2 = 0$$

$$\lambda^2 - (a + c) * \lambda + ac - b^2 = 0$$

$$\lambda = \frac{a + c \pm \sqrt{(a + c)^2 - 4ac + 4b^2}}{2a}$$

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

Solution:

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

$$(a - \lambda)^2 - b^2 = 0$$

$$a - \lambda = \pm b$$

$\lambda = -a - b$ or $\lambda = -a + b$

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 Trace and Eigenvectors [5pts]

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

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

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

Solution:

$$trace(A) = \sum_{n=1}^{N} \lambda_n$$

Since A is symmetric, $A^T = A$.

$$A^T = A = (VAV^T)^T = V^T AV$$
$$= \sum_{n=1}^{N} \lambda_n v_n^T v_n$$
$$= \lambda_1 v_1^T v_1 + \lambda_2 v_2^T v_2 + ... + \lambda_n v_n^T v_n$$
$$= \lambda_1 + \lambda_2 + ... + \lambda_n = \sum_{n=1}^{N} \lambda_n$$

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

Solution:

$$A = A^T = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{\Lambda}\boldsymbol{V}$$
$$V = V^T$$
$$V^T V = V^2$$

1.4 1.3.2 Trace and Eigenvectors **3.5 / 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 Eigenvalue and Eigenvector Calculations [10pts]

Given a matrix

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

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

Solution:

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

$$A - \lambda I = \begin{bmatrix} x - \lambda & 5 \\ 5 & x - \lambda \end{bmatrix}$$

$$(5 - \lambda)^2 - 25 = 0$$

$\lambda = x - 5$ or $\lambda = x + 5$

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

Solution:

1) $\lambda = x - 5$

$$\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} * x = 0$$

$$x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

2) $\lambda = x + 5$

$$\begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix} * x = 0$$

$$x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

1.5 1.3.3 Eigenvalue and Eigenvector Calculations **6 / 10**

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

Error in part a

- **1 pts** Incorrect eigenvalue(s)
- **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 or missing one or two steps

✓ **- 4 pts** *Incomplete or incorrect work for calculating eigenvectors*

- **5 pts** Incorrect eigenvectors and approach
- **7 pts** No work shown
- **7 pts** No attempt or missing solution
- **0.5 pts** The normalized eigenvectors for first lambda are actually representing the same thing.
- **0 pts** Click here to replace this description.

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]

Solution:

X is Bernoulli and Y is Gaussian. $\mu = 0.2, \sigma^2 = 0.96c^2$ for X, as $0.6*c+0.4*-c = 0.2c$, and $E(X^2)-E(x)^2 = c^2 - 0.04c^2 = 0.96c^2$.
$\mu = 0$ and $\sigma^2 = 1$ for Y, as it is normal distribution with given mean and variance.
Z = XY, and X and Y are independent.
mean μ of Z is 0 since $Z = XY$ and $E(XY) = E(X)E(Y)$ for two independent variables.
variance σ^2 of Z is 0 as $E(X^2Y^2) - E(XY)^2 = E(X^2)E(Y^2) - 0 = 0 - 0 = 0$
Therefore, the covariance of Y and Z is going to be:

$$Cov(Y, Z) = E(YZ) - E(Y)E(Z) = E(XY^2) - E(Y)E(Z) = E(XY^2) = E(X)E(Y^2) = 0$$

The answer is 0

2.1 2.1 Covariance **2 / 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 $var(X) = 5$ and $var(Y) = 15$. We do not know $E[X]$ or $E[Y]$. Let $Z = 3X + 2Y$. What is the correlation coefficient $\rho(X, Z) = \frac{cov(X,Z)}{\sqrt{var(X)var(Z)}}$? If applicable, please round your answer to 3 decimal places. [4pts]

Solution:

$$E[Z] = E[3X + 2Y] = 3E[X] + 2E[Y]$$

$$Var(Z) = Var(3X + 2Y) = 9Var(X) + 4Var(Y) + 2Cov(3X, 2Y) = 105 + 12Cov(X, Y)$$

Since X and Y are independent variables, $Var(Z) = 9Var(X) + 4Var(Y) = 105$

$$Cov(X, Y) = 0$$

$$Var(X + 3X + 2Y) = Var(X) + Var(3X) + Var(2Y) + 2Cov(X, 3X + 2Y)$$

$$= 10Var(X) + 4Var(Y) + 6Cov(X, X) + 4Cov(X, Y)$$

$$2Cov(X, Z) = 30 + 0$$

$$Cov(X, Z) = 15$$

$$\rho = \frac{15}{\sqrt{5 * 105}} = 0.655$$

2.2 2.2 Correlation Coefficient **4 / 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 [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. $\min_{x,y} \quad f(x,y) = -4y + xy$) 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).

Solution: Function for maximization:

$$L(x,y) = f(x,y) - \lambda_1 g_1(x,y) - \lambda_2 g_2(x,y)$$
$$= -4y + xy - \lambda_1(2x^2 + y^2 - 12) - \lambda_2(x-1)$$

Function for minimization:

$$L(x,y) = f(x,y) + \lambda_1 g_1(x,y) + \lambda_2 g_2(x,y)$$
$$= -4y + xy + \lambda_1(2x^2 + y^2 - 12) + \lambda_2(x-1)$$

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

Solution:

$$1. \frac{\partial L}{\partial x} = \frac{\partial f(x,y)}{\partial x} - \lambda_1 \frac{\partial g_1(x,y)}{\partial x} - \lambda_2 \frac{\partial g_2(x,y)}{\partial x} = 0$$
$$2. \frac{\partial L}{\partial y} = \frac{\partial f(x,y)}{\partial y} - \lambda_1 \frac{\partial g_1(x,y)}{\partial y} - \lambda_2 \frac{\partial g_2(x,y)}{\partial y} = 0$$
$$3. \lambda_1(2x^2 + y^2 - 12) = 0$$
$$4. \lambda_2(x-1) = 0$$

(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]

Solution:

There are 4 cases - act/act, act/inact, inact/act, inact/inact
1. act/act $\lambda_1 > 0, \lambda_2 > 0$

$$x = 1, y = \pm\sqrt{10}$$
$$y = \sqrt{10}, -3 = 2\sqrt{10}\lambda_1$$
$$\lambda_1 = \frac{-3}{2\sqrt{10}} \text{which is a contradiction}$$

2. act/inact $\lambda_1 > 0, \lambda_2 = 0$

$$y = \pm\sqrt{12 - 2x^2}$$
$$y = \sqrt{12 - 2x^2}$$

$$\sqrt{12-2x^2} - 4x\lambda_1 = 0$$

$$\lambda_1 = \frac{\sqrt{12-2x^2}}{4x}$$

$$-4 + x - 2 * \frac{12-2x^2}{4x} = 0$$

$$x = -1, y = \sqrt{12-2} = \sqrt{10}$$

Solving in a similar way for $y = -12\sqrt{12-2x^2}$,

$$x = \frac{2}{3}, y = \frac{-10}{3}$$

$$answer : (-1, \sqrt{10}), (\frac{2}{3}, \frac{-10}{3})$$

3. inact/act $\lambda_1 = 0, \lambda_2 > 0$
$x = 1, 2x^2 + y^2 - 12 = 0$
$y^2 = 10, y = \pm\sqrt{10}$
$\lambda_2 = \sqrt{10}$

$$answer : (1, \sqrt{10})$$

4. inact/inact $\lambda_1 = 0, \lambda_2 = 0$
Since both λ_1 and λ_2 should be 0, this case causes a contradiction.

(d) List the candidate point(s) (there may be 0, 1, 2, or any number of candidate points) [4pts]

Solution:

3 candidate points: $(-1, \sqrt{10}), (1, \sqrt{10}), (\frac{2}{3}, \frac{-10}{3})$

(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]

Solution:

$$f(x,y) = -4y + xy$$

$$\text{when } (-1, \sqrt{10}) => f(x,y) = -5\sqrt{10}$$

$$\text{when } (1, \sqrt{10}) => f(x,y) = -3\sqrt{10}$$

$$\text{when } (\frac{2}{3}, \frac{-10}{3}) => f(x,y) = \frac{100}{9}$$

$$L(x,y) = -4y + xy - \lambda_1(2x^2 + y^2)$$

$$L_x = y - 5x$$

$$L_y = -4 + x - \frac{5}{2}y$$

$$L_{xy} = -5$$

$$L_{yx} = 1$$

$$H = \begin{bmatrix} -5 & -5 \\ 1 & \frac{-5}{2} \end{bmatrix}$$

$$det(H) > 0, f_(xx) < 0 \text{ at this point it is local max}$$

$$\text{concave at } (\frac{2}{3}, \frac{-10}{3})$$

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).

3.1 Q3 a to e **8 / 15**

- **- 0 pts** Correct
- **- 15 pts** Missing solution
- **- 1 pts** Incorrectly/missing assigned pages

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 or Incorrect answer

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 $$L(x,y)$$
- **1 pts** Incorrect Hessian values
- **1 pts** No comment on convexity
- **2 pts** No attempt or missing solution

(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

3.2 Q3 f (BONUS FOR ALL) **0 / 4**

- **- 0 pts** Correct

✓ **- 4 pts** *Did not attempt/no solution*

- **- 2 pts** No code/incorrect graph
- **- 1 pts** No explanation
- **- 0 pts** Incorrect graph due to issues in previous sections

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

Solution:

6 sided dice for 6 times. It doesn't matter that probability of 1 is different from 2,4,5,6 as we do not have to

$$x_i = \begin{cases} 1 & x = 3 \\ 0 & x \neq 3 \end{cases}$$

$$f(x_i|\theta) = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i}$$

$$L = \prod_{i=1}^{N} p^{x_i}(1-p)^{1-x_i}$$

$$= \theta^0(1-\theta)^{1-0} * \theta^0(1-\theta)^{1-0} * \theta^0(1-\theta)^{1-0} * \theta^0(1-\theta)^{1-0} * \theta^0(1-\theta)^{1-0} * \theta^1$$

$$= (1-\theta)^5\theta$$

Using the MLE can solve this problem. $\frac{\partial L}{\partial \theta} = 0$

$$= (1-\theta)^5 - 5\theta(1-\theta)^4 = 0$$

$$(1-\theta)^4(1-6\theta) = 0$$

$$\theta = \frac{1}{6}$$

4.1 4.1 Discrete Example **7 / 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

No log-likelihood function is mentioned. Only likelihood.

4.2 Weibull distribution [15pts Bonus for Undergrads]

The Weibull distribution is defined as

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

(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 $(x_1, ..., x_n)$, $(X_1, ..., X_n) \sim Weibull(\lambda)$. Here $X_1, ..., X_n$ 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]

4.2 4.2 Weibull Distribution (BONUS FOR UG) **0 / 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.
- **3 pts** Work shown for one term, not the product term
- **5 pts** Inadequate work shown. Incorrect MLE.
- **8 pts** No work. Correct answer.
- **8 pts** Missing steps and incorrect approach
- **9 pts** No work. Incorrect answer.
- **10 pts** Missing submission.

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]

Solution:

$$P(X) = \begin{cases} \frac{5}{16} & x = 0 \\ \frac{11}{16} & x = 1 \end{cases}$$

$$P(Y) = \begin{cases} \frac{1}{4} & y = 0 \\ \frac{3}{4} & y = 1 \end{cases}$$

(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]

Solution:

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

$$= \sum p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

$$= p(0,0) log \frac{p(0,0)}{p(0)p(0)} + p(0,1) log \frac{p(0,1)}{p(0)p(1)}$$

$$+ p(1,0) log \frac{p(1,0)}{p(1)p(0)} + p(1,1) log \frac{p(1,1)}{p(1)p(1)}$$

$$= 0.00450$$

5.1 5.1 Marginal Distribution **6 / 6**

✓ **- 0 pts** *Correct*

- 6 pts Not attempted/No work shown

- 1 pts Incorrectly/missing assigned pages

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]

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}

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

Solution:

$$H(Y) = \sum_{k=1}^{K} P(y=k) log_2 \frac{1}{P(y=k)}$$

$$= P(Y=T) log_2 \frac{1}{P(Y=T)} + P(Y=F) log_2 \frac{1}{P(Y=F)}$$

$$= 0.6 * log_2 \frac{3}{5} + 0.4 * log_2 \frac{5}{2}$$

$$= 0.971$$

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

Solution:

$$H(Y|x_2) = \sum P(x_2, y) log \frac{P(x_2)}{P(x_2, y)}$$

$$= P(x_2 = H, Y = T) log \frac{P(x_2 = H)}{P(x_2 = H, Y = T)}$$

$$+ P(x_2 = H, Y = F) log \frac{P(x_2 = H)}{P(x_2 = H, Y = F)}$$

$$+ P(x_2 = M, Y = T) log \frac{P(x_2 = H)}{P(x_2 = M, Y = T)}$$

$$+ P(x_2 = M, Y = F) log \frac{P(x_2 = H)}{P(x_2 = M, Y = F)}$$

$$+ P(x_2 = L, Y = T) log \frac{P(x_2 = H)}{P(x_2 = L, Y = T)}$$

$$+ P(x_2 = L, Y = F) log \frac{P(x_2 = H)}{P(x_2 = L, Y = F)}$$

$$= 0.761$$

$$H(Y|x_4) = \sum P(x_4, y) log \frac{P(x_4)}{P(x_4, y)}$$

$$= P(x_4 = T, Y = T) log \frac{P(x_4 = H)}{P(x_4 = T, Y = T)}$$

$$+ P(x_4 = T, Y = F) log \frac{P(x_4 = H)}{P(x_4 = T, Y = F)}$$

$$+ P(x_4 = F, Y = T) log \frac{P(x_4 = H)}{P(x_4 = F, Y = T)}$$

$$+ P(x_4 = F, Y = F) log \frac{P(x_4 = H)}{P(x_4 = F, Y = F)}$$

$$= 0.880$$

(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]

Solution:

$$I(x_2, Y) = H(x_2) - H(x_2|Y) = H(Y) - H(Y|x_2)$$

$$= 0.971 - 0.761$$

$$= 0.210$$

$$I(x_4, Y) = H(x_4) - H(x_4|Y) = H(Y) - H(Y|x_4)$$

$$= 0.971 - 0.880$$

$$= 0.091$$

(d) Find joint entropy $H(Y, x_3)$ to at least 3 decimal places. [4pts]

Solution:

$$H(Y, x_3) = \sum P(Y, x_3) log \frac{P(Y)}{P(Y, x_3)}$$

$$= P(Y = T, x_3 = T) log \frac{P(Y = T)}{P(Y = T, x_3 = T)}$$

$$+ P(Y = T, x_3 = F) log \frac{P(Y = T)}{P(Y = T, x_3 = F)}$$

$$+ P(Y = F, x_3 = T) log \frac{P(Y = T)}{P(Y = F, x_3 = T)}$$

$$+ P(Y = F, x_3 = F) log \frac{P(Y = T)}{P(Y = F, x_3 = F)}$$

$$= 0.3 + 0.3 + 0.2 + 0.2$$

$$= 1.000$$

5.2 5.2 Mutual Information and Entropy **16 / 19**

- 0 pts Correct

- 19 pts Not attempted/No work shown

- 1 pts Incorrectly/missing assigned pages

- 5 pts Used log10 instead of log2

error in part a

- 1 pts Minor Error

- 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]

Solution:

$$H(X) = \sum_{k=1}^{K} P(x=k) log_2 \frac{1}{P(x=k)}$$

$$H(X|Y) = \sum_{x \in X, y \in Y} P(x,y) log \frac{P(y)}{P(x,y)}$$

(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]

Solution:

Since X and Y are independent, $P(X|Y) = P(X)$.

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

$$= \sum P(X) log_2 \frac{1}{P(X)} - \sum P(X,Y) log_2 \frac{P(Y)}{P(X,Y)}$$

$$= \sum P(X) log_2 \frac{1}{P(X)} - \sum P(Y) P(X|Y) log_2 \frac{P(Y)}{P(Y)P(X|Y)}$$

$$= \sum P(X) log_2 \frac{1}{P(X)} - \sum P(X) log_2 \frac{1}{P(X)}$$

$$= 0$$

5.3 5.3 Entropy Proofs **10 / 10**

✓ **- 0 pts** *Correct*

- 10 pts Not attempted/No work shown

- 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) 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

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

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]

6 Q6 BONUS FOR ALL 0 / 15

- **0 pts** Correct

✓ **- 15 pts** *Not attempted.*

- **1 pts** Incorrectly/missing assigned pages
- **1 pts** a) Almost correct
- **1 pts** b) minor incorrect
- **4 pts** (a)Minor Correct
- **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

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 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]