Fall 2020 CS4641/CS7641 A Homework 1

Vicente Blat

August 31, 2020

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

Instructions

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

1 Linear Algebra [30pts]

1.1 Determinant and Inverse of Matrix [15pts]

Given a matrix M:

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

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

$$\det M = \begin{vmatrix} r & 6 & 0 \\ 2 & 3 & r \\ 4 & 7 & 3 \end{vmatrix} = r \times \begin{vmatrix} 3 & r \\ 7 & 3 \end{vmatrix} - 6 \times \begin{vmatrix} 2 & r \\ 4 & 3 \end{vmatrix} + 0 \times \begin{vmatrix} 2 & 3 \\ 4 & 7 \end{vmatrix}$$

$$\det M = r \times (3 \times 3 - r \times 7) - 6 \times (2 \times 3 - r \times 4)$$

$$\det M = r \times (9 - 7r) - 6 \times (6 - 4r)$$

$$\det M = 9r - 7r^2 - 36 + 24r$$

$$\boxed{\det M = -7r^2 + 33r - 36}$$

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

The inverse of M will not exist if its determinant is zero. Therefore, to find the values of r for which the inverse will not exist we solve the following equation:

$$0 = -7r^2 + 33r - 36$$

$$r = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$r = \frac{-33 \pm \sqrt{33^2 - 4 \times -7 \times 36}}{2 \times -7}$$

$$r = \frac{-33 \pm \sqrt{1089 - 1008}}{-14}$$

$$r = \frac{-33 \pm \sqrt{81}}{-14}$$

$$r = \frac{-33 \pm 9}{-14}$$

$$r_1 = \frac{-33 + 9}{-14} \qquad r_2 = \frac{-33 - 9}{-14}$$

$$r_1 = \frac{-24}{-14} \qquad r_2 = \frac{-42}{-14}$$

$$r_1 = \frac{12}{7} \qquad r_2 = \frac{21}{7}$$

$$\boxed{r_1 = \frac{12}{7}} \qquad \boxed{r_2 = 3}$$

We find that if r is equal to $\frac{12}{7}$ or 3 the determinant of M is equal to 0, meaning that the matrix will not be linearly independent and therefore its inverse will not exist. In terms of singularity, for both of these values of r the matrix M will be singular since its determinant is equal to zero.

In terms of rank, we would have to substitute r for each of the two values and then reduce the matrix M to row echelon form and count the number of non-zero rows:

For $r_1 = \frac{12}{7}$:

$$\begin{bmatrix} r_1 & 6 & 0 \\ 2 & 3 & r_1 \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{r_1 = \frac{12}{7}} \begin{bmatrix} \frac{12}{7} & 6 & 0 \\ 2 & 3 & \frac{12}{7} \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_1 = R_1 / \frac{12}{7}} \begin{bmatrix} 1 & \frac{7}{2} & 0 \\ 2 & 3 & \frac{12}{7} \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_2 = R_2 - 2R_1}$$

$$\begin{bmatrix} 1 & \frac{7}{2} & 0 \\ 0 & -4 & \frac{12}{7} \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_3 = R_3 - 4R_1} \begin{bmatrix} 1 & \frac{7}{2} & 0 \\ 0 & -4 & \frac{12}{7} \\ 0 & -7 & 3 \end{bmatrix} \xrightarrow{R_2 = R_2 / -4} \begin{bmatrix} 1 & \frac{7}{2} & 0 \\ 0 & 1 & \frac{-3}{7} \\ 0 & -7 & 3 \end{bmatrix} \xrightarrow{R_1 = R_1 - \frac{7}{2} R_2}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & \frac{7}{2} \\ 0 & -7 & 3 \end{bmatrix} \xrightarrow{R_3 = R_3 + 7R_2} \begin{bmatrix} 1 & 0 & \frac{3}{2} \\ 0 & 1 & \frac{-3}{7} \\ 0 & 0 & 0 \end{bmatrix}$$

The rank of matrix M when $r_1 = \frac{12}{7}$ is equal to 2 because after reducing the matrix to row echelon form there are only 2 non-zero rows.

For $r_2 = 3$:

$$\begin{bmatrix} r_2 & 6 & 0 \\ 2 & 3 & r_2 \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{r_2 = 3} \begin{bmatrix} 3 & 6 & 0 \\ 2 & 3 & 3 \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_1 = R_1 / 3} \begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & 3 \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_2 = R_2 - 2R_1} \begin{bmatrix} 1 & 2 & 0 \\ 0 & -1 & 3 \\ 4 & 7 & 3 \end{bmatrix} \xrightarrow{R_3 = R_3 - 4R_1}$$

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & -1 & 3 \\ 0 & -1 & 3 \end{bmatrix} \xrightarrow{R_2 = R_2 / -1} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & -3 \\ 0 & -1 & 3 \end{bmatrix} \xrightarrow{R_1 = R_1 - 2R_2} \begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & -3 \\ 0 & -1 & 3 \end{bmatrix} \xrightarrow{R_3 = R_3 + R_2}$$

$$\begin{bmatrix} 1 & 0 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

The rank of matrix M when $r_1 = 3$ is equal to 2 because after reducing the matrix to row echelon form there are only 2 non-zero rows.

(c) Calculate M^{-1} by hand for $r = 4$. [5pts]

$$M = \left(\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_1=R_1/4} \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2=R_2-2R_1}$$

$$M = \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 4 & \frac{-1}{2} & 1 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2=R_2/4} \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 1 & \frac{-1}{8} & \frac{1}{4} & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2 \Longleftrightarrow R_3}$$

$$M = \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{-1}{8} & \frac{1}{4} & 0 \end{array}\right) \xrightarrow{R_2=R_2-4R_1} \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 1 & \frac{-1}{8} & \frac{1}{4} & 0 \end{array}\right) \xrightarrow{R_2=R_2-3R_3}$$

$$M = \left(\begin{array}{ccc|ccc} 1 & 3/2 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 1 & 0 & \frac{-5}{8} & \frac{-3}{4} & 1 \\ 0 & 0 & 1 & \frac{-1}{8} & \frac{1}{4} & 0 \end{array}\right) \xrightarrow{R_1=R_1-\frac{3}{2}R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{19}{16} & \frac{9}{8} & \frac{-3}{2} \\ 0 & 1 & 0 & \frac{-5}{8} & \frac{-3}{4} & 1 \\ 0 & 0 & 1 & \frac{-1}{8} & \frac{1}{4} & 0 \end{array}\right)$$

Therefore, the inverse of M is:

$$M^{-1} = \begin{bmatrix} \frac{19}{16} & \frac{9}{8} & \frac{-3}{2} \\ \frac{-5}{8} & \frac{-3}{4} & 1 \\ \frac{-1}{8} & \frac{1}{4} & 0 \end{bmatrix}$$

(**Hint 1:** Please double check your answer and make sure $MM^{-1} = I$)

$M \times M^{-1} = I$

$$\begin{bmatrix} 4 & 6 & 0 \\ 2 & 3 & 4 \\ 4 & 7 & 3 \end{bmatrix} \times \begin{bmatrix} \frac{19}{16} & \frac{9}{8} & \frac{-3}{2} \\ \frac{-5}{8} & \frac{-3}{4} & 1 \\ \frac{-1}{8} & \frac{1}{4} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} (4 \times \frac{19}{16} + 6 \times \frac{-5}{8} + 0 \times \frac{-1}{8}) & (4 \times \frac{9}{8} + 6 \times \frac{-3}{4} + 0 \times \frac{1}{4}) & (4 \times \frac{-3}{2} + 6 \times 1 + 0 \times 0) \\ (2 \times \frac{19}{16} + 3 \times \frac{-5}{8} + 4 \times \frac{-1}{8}) & (2 \times \frac{9}{8} + 3 \times \frac{-3}{4} + 4 \times \frac{1}{4}) & (2 \times \frac{-3}{2} + 3 \times 1 + 4 \times 0) \\ (4 \times \frac{19}{16} + 7 \times \frac{-5}{8} + 3 \times \frac{-1}{8}) & (4 \times \frac{9}{8} + 7 \times \frac{-3}{4} + 3 \times \frac{1}{4}) & (4 \times \frac{-3}{2} + 7 \times 1 + 3 \times 0) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} (\frac{19}{4} - \frac{15}{4}) & (\frac{9}{2} - \frac{9}{2}) & (-6 + 6) \\ (\frac{19}{8} - \frac{15}{8} - \frac{1}{2}) & (\frac{9}{4} - \frac{9}{4} + 1) & (-3 + 3) \\ (\frac{19}{4} - \frac{35}{8} - \frac{3}{8}) & (\frac{9}{2} - \frac{21}{4} + \frac{3}{4}) & (-6 + 7) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

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

$$\det M^{-1} = \begin{vmatrix} \frac{19}{16} & \frac{9}{8} & \frac{-3}{2} \\ \frac{-5}{8} & \frac{-3}{4} & 1 \\ \frac{-1}{8} & \frac{1}{4} & 0 \end{vmatrix} = \frac{19}{16} \times \begin{vmatrix} \frac{-3}{4} & 1 \\ \frac{1}{4} & 0 \end{vmatrix} - \frac{9}{8} \times \begin{vmatrix} \frac{-5}{8} & 1 \\ \frac{-1}{8} & 0 \end{vmatrix} + \frac{-3}{2} \times \begin{vmatrix} \frac{-5}{8} & \frac{-3}{4} \\ \frac{-1}{8} & \frac{1}{4} \end{vmatrix}$$

$$\det M^{-1} = \frac{19}{16} \times (\frac{-3}{4} \times 0 - 1 \times \frac{1}{4}) - \frac{9}{8} \times (\frac{-5}{8} \times 0 - 1 \times \frac{-1}{8}) - \frac{3}{2} \times (\frac{-5}{8} \times \frac{1}{4} - \frac{-3}{4} \times \frac{-1}{8})$$

$$\det M^{-1} = (\frac{19}{16} \times \frac{-1}{4}) - (\frac{9}{8} \times \frac{1}{8}) - (\frac{3}{2} \times \frac{-1}{4})$$

$$\boxed{\det M^{-1} = -\frac{1}{16}}$$

1.2 Characteristic Equation [5pts]

Consider the eigenvalue problem:

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

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

$|A - \lambda I| = 0$ is true when $(A - \lambda I)$ is **not** an invertible matrix, which happens when either $(A - \lambda I) = 0$ or when the matrix is linearly dependent. We can prove this is true given the initial equality:

$$\begin{aligned}
Ax &= \lambda x \\
Ax - \lambda x &= 0 \\
(A - \lambda)x &= 0, && \text{Multiply by identity matrix I} \\
(AI - \lambda I)x &= 0, && \text{Note that } AI = A \\
(A - \lambda I)x &= 0
\end{aligned}$$

Given that $x \neq 0$, the set of vectors in $A - \lambda I$ must be linearly dependent or $A - \lambda I = 0$. Either way, the determinant $|A - \lambda I|$ will be equal to zero.

1.3 Eigenvalues and Eigenvectors [10pts]

Given a matrix A:

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

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

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

$$\begin{vmatrix} x-\lambda & 3 \\ 1 & x-\lambda \end{vmatrix} = 0$$

$$(x-\lambda)^2 - 3 = 0$$

$$(x-\lambda)^2 = 3$$

$$x - \lambda_1 = \sqrt{3} \qquad x - \lambda_2 = -\sqrt{3}$$

$$\boxed{x - \sqrt{3} = \lambda_1} \qquad \boxed{x + \sqrt{3} = \lambda_2}$$

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

$$Av = \lambda v$$

$$(A - \lambda)v = 0$$

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

$$\left(\begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$$

For $\lambda_1 = x - \sqrt{3}$

$$\left(\begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} - \begin{bmatrix} x - \sqrt{3} & 0 \\ 0 & x - \sqrt{3} \end{bmatrix}\right)\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 0$$

$$\left(\begin{bmatrix} \sqrt{3} & 3 \\ 1 & \sqrt{3} \end{bmatrix}\right)\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 0$$

Find the reduced echelon form of $(A - \lambda I)$

$$\left(\begin{array}{cc|c} \sqrt{3} & 3 & 0 \\ 1 & \sqrt{3} & 0 \end{array}\right) \xrightarrow{R_1 = R_1/\sqrt{3}} \left(\begin{array}{cc|c} 1 & \sqrt{3} & 0 \\ 1 & \sqrt{3} & 0 \end{array}\right) \xrightarrow{R_2 = R_2 - R_1} \left(\begin{array}{cc|c} 1 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{array}\right)$$

We find that:

$$v_{11} + \sqrt{3}v_{12} = 0$$

$$v_{11} = -\sqrt{3}v_{12}$$

Let $v_{12} = 1$, then $v_{11} = -\sqrt{3}$. For the eigenvalue $\lambda_1 = x - \sqrt{3}$, the eigenvector is:

$$v_1 = \begin{bmatrix} -\sqrt{3} \\ 1 \end{bmatrix} \rightarrow v_{1n} = \frac{v_1}{||v_1||} \rightarrow v_{1n} = \frac{\begin{bmatrix} -\sqrt{3} \\ 1 \end{bmatrix}}{\sqrt{(-\sqrt{3})^2 + 1}} \rightarrow \boxed{v_{1n} = \frac{1}{2}\begin{bmatrix} -\sqrt{3} \\ 1 \end{bmatrix}}$$

For $\lambda_2 = x + \sqrt{3}$

$$\left(\begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} - \begin{bmatrix} x+\sqrt{3} & 0 \\ 0 & x+\sqrt{3} \end{bmatrix}\right)\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = 0$$

$$\left(\begin{bmatrix} -\sqrt{3} & 3 \\ 1 & -\sqrt{3} \end{bmatrix}\right)\begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix} = 0$$

Find the reduced echelon form of $(A - \lambda I)$

$$\left(\begin{array}{cc|c} -\sqrt{3} & 3 & 0 \\ 1 & -\sqrt{3} & 0 \end{array}\right) \xrightarrow{R_1 = R_1/-\sqrt{3}} \left(\begin{array}{cc|c} 1 & -\sqrt{3} & 0 \\ 1 & -\sqrt{3} & 0 \end{array}\right) \xrightarrow{R_2 = R_2 - R_1} \left(\begin{array}{cc|c} 1 & -\sqrt{3} & 0 \\ 0 & 0 & 0 \end{array}\right)$$

We find that:

$$v_{21} - \sqrt{3}v_{22} = 0$$

$$v_{21} = \sqrt{3}v_{22}$$

Let $v_{22} = 1$, then $v_{21} = \sqrt{3}$. For the eigenvalue $\lambda_2 = x + \sqrt{3}$, the eigenvector is:

$$v_2 = \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} \rightarrow v_{2n} = \frac{v_2}{||v_2||} \rightarrow v_{2n} = \frac{\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}}{\sqrt{(\sqrt{3})^2 + 1}} \rightarrow \boxed{v_{2n} = \frac{1}{2}\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}}$$

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

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

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

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

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

$$P(Z \geq z) = P(XY \geq z)$$

$$= P(XY \geq z | X = c)P(X = c) + P(XY \geq z | X = c)P(X = -c)$$

$$= \frac{1}{2}P(XY \geq z | X = c) + \frac{1}{2}P(XY \geq z | X = -c)$$

$$= \frac{1}{2}P(cY \geq z) + \frac{1}{2}P(-cY \geq z)$$

$$= \frac{1}{2}P(cY \geq z) + \frac{1}{2}P(cY \leq -z)$$

Due to symmetry $P(cY \geq z) = P(cY \leq -z)$, therefore:

$$P(Z \geq z) = P(cY \geq z)$$

Then, we can say that $Z = cY$ or $Z = \sum_{n=1}^{c} Y$. And according to one of the properties of Gaussian distributions the sum of Gaussian distributions results in another Gaussian distribution we can claim that Z does follow a Gaussian distribution.

The expected value of Z is then:

$$E[Z] = E[cY]$$

$$= cE[Y], \quad \text{Given that } E[Y] = 0$$

$$E[Z] = 0$$

The variance of Z is then:

$$Var[Z] = Var[cY]$$

$$= c^2 Var[Y], \quad \text{Given that } Var[Y] = 1$$

$$Var[Z] = c^2$$

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

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

$$E[YZ] - E[Y]E[Z] = 0$$

$$E[cY^2] - E[Y]E[cY] = 0, \quad \text{Given that } E[Y] = 0$$

$$cE[Y^2] = 0$$

Therefore, in order for Y and Z to be uncorrelated we must choose c to be: $c = 0$

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

If Y and Z are independent, then it must be true that $P(Y|Z) = P(Y)$. Let us try to prove or disprove this statement with an example:

$$P(Y|Z) = P(Y)$$

$$P(-1 \leq y \leq 0 \mid 2c \leq z \leq 3c) = P(-1 \leq y \leq 0)$$

$$P(-1 \leq y \leq 0 \mid 2c \leq cy \leq 3c) = P(-1 \leq y \leq 0)$$

$$P(-1 \leq y \leq 0 \mid 2 \leq y \leq 3) = P(-1 \leq y \leq 0)$$

This statement is a contradiction because: $P(-1 \leq y \leq 0 \mid 2 \leq y \leq 3)$ is an impossible event, therefore we find that $P(Y|Z) \neq P(Y)$ and thus Y and Z are NOT independent.

3 Optimization [15 pts]

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

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

(a) Specify the Legrange function [2 pts]

$$\mathcal{L}(x,y) = 2x^2 + 3xy - \lambda_1(\frac{1}{2}x^2 + y - 4) - \lambda_2(-y+2), \text{where: } \lambda_1, \lambda_2 > 0$$

(b) List the KKT conditions [2 pts]

(i) $\frac{\partial \mathcal{L}}{\partial x} = 4x + 3y - \lambda_1 x = 0$

(ii) $\frac{\partial \mathcal{L}}{\partial y} = 3x - \lambda_1 + \lambda_2 = 0$

(iii) $\lambda_1(\frac{1}{2}x^2 + y - 4) = 0$

(iv) $\lambda_2(-y + 2) = 0$

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

Case 1: constraint 1 and constraint 2 are active.

$$\frac{1}{2}x^2 + y = 4,\ \lambda_1 > 0 \qquad y = 2,\ \lambda_2 > 0$$

$$\begin{aligned} \frac{1}{2}x^2 + (2) &= 4 \\ \frac{1}{2}x^2 &= 2 \\ x^2 &= 4 \\ x &= \pm 2 \end{aligned}$$

For $x = 2, y = 2$

$$\begin{aligned} (i)\ 4x + 3y - \lambda_1 x &= 0 \\ 8 + 6 - 2\lambda_1 &= 0 \\ -2\lambda_1 &= -14 \\ &\boxed{\lambda_1 = 7 > 0\ \textit{acceptable}} \end{aligned}$$

$$\begin{aligned} (ii)\ 3x - \lambda_1 + \lambda_2 &= 0 \\ 6 - 7 + \lambda_2 &= 0 \\ -1 + \lambda_2 &= 0 \\ &\boxed{\lambda_2 = 1 > 0\ \textit{acceptable}} \end{aligned}$$

For $x = -2, y = 2$

$$\begin{aligned} (i)\ 4x + 3y - \lambda_1 x &= 0 \\ -8 + 6 + 2\lambda_1 &= 0 \\ 2\lambda_1 &= 2 \\ &\boxed{\lambda_1 = 1 > 0\ \textit{acceptable}} \end{aligned}$$

$$\begin{aligned} (ii)\ 3x - \lambda_1 + \lambda_2 &= 0 \\ -6 - 1 + \lambda_2 &= 0 \\ -7 + \lambda_2 &= 0 \\ &\boxed{\lambda_2 = 7 > 0\ \textit{acceptable}} \end{aligned}$$

Case 2: constraint 1 is active, constraint 2 is inactive.

$$\frac{1}{2}x^2 + y = 4,\ \lambda_1 > 0 \qquad y > 2,\ \lambda_2 = 0$$

$$\begin{aligned} (ii)\ 3x - \lambda_1 + \lambda_2 &= 0 \\ 3x - \lambda_1 &= 0 \end{aligned} \qquad \begin{aligned} 4 &= \frac{1}{2}x^2 + y \\ y &= 4 - \frac{1}{2}x^2 \end{aligned}$$

$$
\begin{aligned}
(i)\ 4x + 3y - \lambda_1 x &= 0 \\
4x + 3(4 - \frac{1}{2}x^2) - (3x)x &= 0 \\
4x + 12 - \frac{3}{2}x^2 - 3x^2 &= 0 \\
-\frac{9}{2}x^2 + 4x + 12 &= 0 \\
-9x^2 + 8x + 24 &= 0
\end{aligned}
\qquad
\begin{aligned}
x &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-8 \pm \sqrt{8^2 - 4(-9)(24)}}{2(-9)} \\
x &= \frac{-8 \pm \sqrt{64 + 864}}{-18} \\
x &= \frac{-8 \pm 4\sqrt{58}}{-18}
\end{aligned}
$$

$$x = \frac{4 \pm 2\sqrt{58}}{9}$$

For $x = \frac{4+2\sqrt{58}}{9}$

$$
\begin{aligned}
y &= 4 - \frac{1}{2}x^2 \\
y &= 4 - \frac{1}{2}(\frac{4 + 2\sqrt{58}}{9})^2 \\
y &= 4 - \frac{16 + 16\sqrt{58} + 232}{(2)(9)}
\end{aligned}
$$

$$\boxed{y = 4 - \frac{8 + 8\sqrt{58} + 116}{9} \approx 1.71696 < 2 \textit{ not acceptable}}$$

For $x = \frac{4-2\sqrt{58}}{9}$

$$
\begin{aligned}
\lambda_1 &= 3x \\
\lambda_1 &= 3(\frac{4 - 2\sqrt{58}}{9})
\end{aligned}
$$

$$\boxed{\lambda_1 = \frac{4 - 2\sqrt{58}}{3} \approx -3.74385 < 0 \textit{ not acceptable}}$$

Case 3: Constraint 1 is inactive, constraint 2 is active.

$$\frac{1}{2}x^2 + y < 4,\ \lambda_1 = 0 \qquad y = 2,\ \lambda_2 > 0$$

$$\begin{aligned} (i)\ 4x + 3y - \lambda_1 x &= 0 \\ 4x + 6 &= 0 \\ x &= -\frac{6}{4} \\ x &= -\frac{3}{2} \end{aligned} \qquad \begin{aligned} (ii)\ 3x - \lambda_1 + \lambda_2 &= 0 \\ 3x + \lambda_2 &= 0 \\ \frac{-9}{2} + \lambda_2 &= 0 \\ \boxed{\lambda_2 = \frac{9}{2} > 0\ \textit{acceptable}} \end{aligned}$$

Case 4: Constraint 1 is inactive, constraint 2 is inactive.

$$\frac{1}{2}x^2 + y < 4,\ \lambda_1 = 0 \qquad y > 2,\ \lambda_2 = 0$$

$$\begin{aligned} (ii)\ 3x - \lambda_1 + \lambda_2 &= 0 \\ \boxed{3x = 0\ \textit{not acceptable}} \end{aligned}$$

(d) List all candidate points [4 pts]

- Candidate 1: $(x, y) = (2, 2)\ , with\ (\lambda_1, \lambda_2) = (7, 1)$
- Candidate 2: $(x, y) = (-2, 2)\ , with\ (\lambda_1, \lambda_2) = (1, 7)$
- Candidate 3: $(x, y) = (-\frac{3}{2}, 2)\ , with\ (\lambda_1, \lambda_2) = (0, \frac{9}{2})$

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

Maximality:

Candidate 1:

$$f(2,2) = 2(2)^2 + 3(2)(2)$$

$$f(2,2) = 20$$

Candidate 2:

$$f(-2,2) = 2(-2)^2 + 3(-2)(2)$$

$$f(-2,2) = -4$$

Candidate 3:

$$f(-\frac{3}{2},2) = 2(-\frac{3}{2})^2 + 3(-\frac{3}{2})(2)$$

$$f(-\frac{3}{2},2) = \frac{9}{2} - 9$$

$$f(-\frac{3}{2},2) = -\frac{9}{2}$$

We find that $f(2,2) = 20$ is the maximum value, we now check this candidate for sufficiency: $\mathcal{L}(x,y),\ with\ \lambda_1 = 7\ and\ \lambda_2 = 1$

$$\mathcal{L}(x,y) = 2x^2 + 3xy - (7)(\frac{1}{2}x^2 + y - 4) - (1)(-y+2)$$

$$\mathcal{L}(x,y) = 2x^2 + 3xy - \frac{7}{2}x^2 - 7y + 28 + y - 2$$

$$\mathcal{L}(x,y) = -\frac{3}{2}x^2 + 3xy - 6y + 26$$

We find that this candidate point is also sufficient since the Legrange function for $\lambda_1 = 7$ and $\lambda_2 = 1$ is a concave function. Therefore, the maximum and sufficient candidate point is:

$(x,y) = (2,2)\ ,with\ (\lambda_1, \lambda_2) = (7,1)$

4 Maximum Likelihood [10 + 25 pts]

4.1 Discrete Example [10 pts]

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

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

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

Since we are dealing with coin flips, we can assume that both A and B are independent and identically distributed random variables with Bernoulli distribution objective functions:

$f(A; p_A) = (p_A)^{a_i}(1 - p_A)^{(1-a_i)}$ Given that $a_i \in \{0, 1\}$ or $\{tail, head\}$, and $p_A = \theta$

$f(B; p_B) = (p_B)^{b_i}(1 - p_B)^{(1-b_i)}$ Given that $b_i \in \{0, 1\}$ or $\{tail, head\}$, and $p_B = 2\theta$

Because A and B are independent of each other we can say that the likelihood of the result is:

$$\begin{aligned} L(p) &= L(p_A)L(p_B) \\ &= \prod_{i=1}^{n_A} f(A; p_A) \prod_{j=1}^{n_B} f(B; p_B) \text{ ,where } n_A = 7 \text{ and } n_B = 3 \\ &= \prod_{i=1}^{7}[\theta^{a_i}(1-\theta)^{(1-a_i)}] \prod_{j=1}^{3}[(2\theta)^{b_j}(1-2\theta)^{(1-b_j)}] \\ &= [\theta^{\sum_{i=1}^{7} a_i}(1-\theta)^{\sum_{i=1}^{7}(1-a_i)}][(2\theta)^{\sum_{j=1}^{3} b_j}(1-2\theta)^{\sum_{j=1}^{3}(1-b_j)}] \end{aligned}$$

From the sequence given we find that:

$\sum_{i=1}^{7} a_i = \{0+0+0+0+0+1+1\} = 2$

$\sum_{i=1}^{7}(1-a_i) = \{1+1+1+1+1+0+0\} = 5$

$\sum_{j=1}^{3} b_j = \{1+1+1\} = 3$

$\sum_{j=1}^{3}(1-b_j) = \{0+0+0\} = 0$

Therefore, the likelihood of the result given θ is:

$$L(\theta) = [\theta^2(1-\theta)^5][(2\theta)^3(1-2\theta)^0]$$

$$= 8\theta^5(1-\theta)^5$$

$$\boxed{= \log(8) + 5\log(\theta) + 5\log(1-\theta)}$$

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

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \implies \frac{5}{\theta} - \frac{5}{1-\theta} = 0 \implies \frac{1}{\theta} = \frac{1}{1-\theta}$$

$$\frac{1-\theta}{\theta} = 1 \implies \frac{1}{\theta} - 1 = 1 \implies \frac{1}{\theta} = 2 \implies \qquad \boxed{\theta = \frac{1}{2}}$$

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

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

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

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

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

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

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

$$Var[X_n] = E[(X_n - E[X])^2]$$

$$E[X_n] = E[\theta t_n^3 + Z_n]$$

$$= E[(\theta t_n^3 + Z_n - \theta t_n^3)^2]$$

$$= E[\theta t_n^3] + E[Z_n]$$

$$= E[(Z_n)^2]$$

$$= t_n^3 E[\theta] + 0$$

$$= Var[Z_n] + (E[Z_n])^2$$

$$\boxed{E[X_n] = \theta t_n^3}$$

$$= \sigma^2 + 0$$

$$\boxed{Var[X_n] = \sigma^2}$$

We can confirm that X_n follows a Gaussian distribution since it is equal to the Gaussian distribution Z_n shifted by θt_n^3. And according to one of the properties of Normal distributions, a Normal distribution plus a constant is a new Normal distribution.

$$L(\theta; x_1, ..., x_N) = \prod_{n=1}^{N} f_{Xn}(x_n; \theta)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_n - \theta t_n^3)^2}{2\sigma^2}}$$

$$= (2\pi\sigma^2)^{\frac{-N}{2}} e^{(-\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \theta t_n^3)^2)}$$

$$\boxed{L(\theta; x_1, ..., x_N) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \theta t_n^3)^2}$$

(b) Compute the MLE for θ. [6pts]

$$\frac{\partial}{\partial\theta}L(\theta; x_1, ..., x_N) = 0$$

$$\left(-\frac{2}{2\sigma^2}\sum_{n=1}^{N}(x_n - \theta t_n^3)\right)(-t_n^3) = 0$$

$$\frac{t_n^3}{\sigma^2}\sum_{n=1}^{N}(x_n - \theta t_n^3) = 0$$

$$\sum_{n=1}^{N}(x_n) - N\theta t_n^3 = 0$$

$$N\theta t_n^3 = \sum_{n=1}^{N}(x_n)$$

$$\boxed{\theta = \frac{1}{Nt_n^3}\sum_{n=1}^{N}(x_n)}$$

4.3 Bonus for undergrads [10 pts]

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

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

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

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

$$f_X(X_i \leq x|\alpha, \beta) = \frac{d}{dx} P(X_i \leq x|\alpha, \beta) = \begin{cases} \left(\frac{\alpha}{\beta^\alpha}\right) x^{\alpha-1}, & 0 \leq x \leq \beta \\ \\ 0, & \text{everywhere else} \end{cases}$$

(b) Find the MLEs of α and β. [6pts]

$$L(\alpha,\beta;x_1,...,x_N) = \prod_{n=1}^{N} f_{Xn}(x_n;\alpha)$$

$$= \prod_{n=1}^{N} \frac{\alpha}{\beta^\alpha} x_n^{(\alpha-1)}$$

$$= \sum_{n=1}^{N} \left(\log(\frac{\alpha}{\beta^\alpha} x_n^{(\alpha-1)}) \right)$$

$$= \sum_{n=1}^{N} \left(\log(\frac{\alpha}{\beta^\alpha}) + \log(x_n^{(\alpha-1)}) \right)$$

$$L(\alpha,\beta;x_1,...,x_N) = \sum_{n=1}^{N} (\log(\alpha) - \alpha\log(\beta) + (\alpha-1)\log(x_n))$$

$$\frac{\partial}{\partial\alpha} L(\alpha;x_1,...,x_N) = 0$$

$$\sum_{n=1}^{N} (\frac{1}{\alpha} - \log(\beta) + \log(x_n)) = 0$$

$$N\log(\beta) - \sum_{n=1}^{N}\log(x_n) = \frac{N}{\alpha}$$

$$\boxed{\frac{N}{N\log(\beta) - \sum_{n=1}^{N}\log(x_n)} = \alpha}$$

$$\frac{\partial}{\partial\beta} L(\alpha;x_1,...,x_N) = 0$$

$$\sum_{n=1}^{N} (-\frac{\alpha}{\beta}) = 0$$

$$-N\frac{\alpha}{\beta} = 0$$

$$\boxed{\beta = 0}$$

5 Information Theory [32pts]

5.1 Marginal Distribution [6pts]

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

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

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

For $X = 0$:

$$P(X=0) = \sum_{Y} P(X=0, Y)$$

$$= P(X=0, Y=1) + P(X=0, Y=2)$$

$$= \frac{1}{3} + \frac{1}{3}$$

$$\boxed{P(X=0) = \frac{2}{3}}$$

For $X = 1$:

$$P(X=1) = \sum_{Y} P(X=1, Y)$$

$$= P(X=1, Y=1) + P(X=1, Y=2)$$

$$= 0 + \frac{1}{3}$$

$$\boxed{P(X=1) = \frac{1}{3}}$$

For $Y = 1$:

$$P(Y = 1) = \sum_X P(Y = 1, X)$$

$$= P(Y = 1, X = 0) + P(Y = 1, X = 1)$$

$$= \frac{1}{3} + 0$$

$$\boxed{P(Y = 1) = \frac{1}{3}}$$

For $Y = 2$:

$$P(Y = 2) = \sum_X P(Y = 2, X)$$

$$= P(Y = 2, X = 0) + P(Y = 2, X = 1)$$

$$= \frac{1}{3} + \frac{1}{3}$$

$$\boxed{P(Y = 2) = \frac{2}{3}}$$

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

$$I(X,Y) = \sum_X \sum_Y P(X,Y) \log_2 \left(\frac{P(X,Y)}{P(X)P(Y)} \right)$$

$$= \sum_X \left[P(X, Y=1) \log_2 \left(\frac{P(X,Y=1)}{P(X)P(Y=1)} \right) + P(X,Y=2) \log_2 \left(\frac{P(X,Y=2)}{P(X)P(Y=2)} \right) \right]$$

$$= P(X=0, Y=1) \log_2 \left(\frac{P(X=0,Y=1)}{P(X=0)P(Y=1)} \right) + P(X=1,Y=1) \log_2 \left(\frac{P(X=1,Y=1)}{P(X=1)P(Y=1)} \right)$$

$$+ P(X=0, Y=2) \log_2 \left(\frac{P(X=0,Y=2)}{P(X=0)P(Y=2)} \right) + P(X=1,Y=2) \log_2 \left(\frac{P(X=1,Y=2)}{P(X=1)P(Y=2)} \right)$$

$$= \frac{1}{3} \log_2 \left(\frac{\frac{1}{3}}{\frac{2}{9}} \right) + 0 \log_2 \left(\frac{0}{\frac{1}{9}} \right) + \frac{1}{3} \log_2 \left(\frac{\frac{1}{3}}{\frac{4}{9}} \right) + \frac{1}{3} \log_2 \left(\frac{\frac{1}{3}}{\frac{2}{9}} \right)$$

$$= \frac{1}{3} \left(2 \log_2 \left(\frac{3}{2} \right) + \log_2 \left(\frac{3}{4} \right) \right)$$

$$= \frac{1}{3} (2 \log_2(3) - 2 \log_2(2) + \log_2(3) - \log_2(4))$$

$$= \frac{1}{3} (3 \log_2(3) - 4)$$

$$= \log_2(3) - \frac{4}{3}$$

$$\boxed{I(X,Y) = 0.25162}$$

5.2 Mutual Information and Entropy [19pts]

Given a dataset as below.

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

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

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

$$Y = \{yes, no\} = \{1, 0\}$$

$$P(y = 1) = \frac{8}{14} = \frac{4}{7} \qquad P(y = 0) = \frac{6}{14} = \frac{3}{7}$$

$$H(Y) = \sum_{i=1}^{n} P(Y = y_i) I(Y)$$

$$= \sum_{i=1}^{n} P(Y = y_i) \log_2 \left(\frac{1}{P(y_i)} \right)$$

$$= -\sum_{i=1}^{n} P(Y = y_i) \log_2 P(Y = y_i)$$

$$= -P(y = 1) \log_2 P(y = 1) - P(y = 0) \log_2 P(y = 0)$$

$$= -\frac{4}{7} \log_2 \left(\frac{4}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right)$$

$$= -\frac{1}{7} [4 \log_2(4) - 4 \log_2(7) + 3 \log_2(3) - 3 \log_2(7)]$$

$$= -\frac{1}{7} [8 + 3 \log_2(3) - 7 \log_2(7)]$$

$$= -\frac{8}{7} - \frac{3}{7} \log_2(3) + \log_2(7)$$

$$= -1.14286 - 0.67926 + 2.80735$$

$$\boxed{H(Y) = 0.98523}$$

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

For $H(Y|x_1)$:

$Y = \{yes, no\} = \{1, 0\}$ $\qquad$ $x_1 = \{young, middleage, senior\} = \{g, m, s\}$

$$P(x_1 = g) = \frac{5}{14} \qquad P(x_1 = m) = \frac{4}{14} = \frac{2}{7} \qquad P(x_1 = s) = \frac{5}{14}$$

$$P(x_1 = g, y = 1) = \frac{1}{14} \qquad P(x_1 = g, y = 0) = \frac{4}{14} = \frac{2}{7}$$

$$P(x_1 = m, y = 1) = \frac{4}{14} = \frac{2}{7} \qquad P(x_1 = m, y = 0) = 0$$

$$P(x_1 = s, y = 1) = \frac{3}{14} \qquad P(x_1 = s, y = 0) = \frac{2}{14} = \frac{1}{7}$$

$$H(Y|x_1) = \sum_{x_1 \epsilon X} \sum_{y \epsilon Y} P(x_1, y) \log_2 \left(\frac{P(x_1)}{P(x_1, y)} \right)$$

$$= P(x_1 = g, y = 1) \log_2 \left(\frac{P(x_1 = g)}{P(x_1 = g, y = 1)} \right) + P(x_1 = m, y = 1) \log_2 \left(\frac{P(x_1 = m)}{P(x_1 = m, y = 1)} \right)$$

$$+ P(x_1 = s, y = 1) \log_2 \left(\frac{P(x_1 = s)}{P(x_1 = s, y = 1)} \right) + P(x_1 = g, y = 0) \log_2 \left(\frac{P(x_1 = g)}{P(x_1 = g, y = 0)} \right)$$

$$+ P(x_1 = m, y = 0) \log_2 \left(\frac{P(x_1 = m)}{P(x_1 = m, y = 0)} \right) + P(x_1 = s, y = 0) \log_2 \left(\frac{P(x_1 = s)}{P(x_1 = s, y = 0)} \right)$$

$$= \frac{1}{14} \log_2 \left(\frac{\frac{5}{14}}{\frac{1}{14}} \right) + \frac{2}{7} \log_2 \left(\frac{\frac{2}{7}}{\frac{2}{7}} \right) + \frac{3}{14} \log_2 \left(\frac{\frac{5}{14}}{\frac{3}{14}} \right) + \frac{2}{7} \log_2 \left(\frac{\frac{5}{14}}{\frac{2}{7}} \right) + 0 + \frac{1}{7} \log_2 \left(\frac{\frac{5}{14}}{\frac{1}{7}} \right)$$

$$= \frac{1}{14} \log_2(5) + \frac{2}{7} \log_2(1) + \frac{3}{14} \log_2 \left(\frac{5}{3} \right) + \frac{2}{7} \log_2 \left(\frac{5}{4} \right) + \frac{1}{7} \log_2 \left(\frac{5}{2} \right)$$

$$= 0.16585 + 0 + 0.15792 + 0.09197 + 0.18884$$

$\boxed{H(Y|x_1) = 0.60458}$

For $H(Y|x_4)$:

$$Y = \{yes, no\} = \{1, 0\} \qquad x_4 = \{yes, no\} = \{1, 0\}$$

$$P(x_4 = 1) = \frac{8}{14} = \frac{4}{7} \qquad P(x_4 = 0) = \frac{6}{14} = \frac{3}{7}$$

$$P(x_4 = 1, y = 1) = \frac{5}{14} \qquad P(x_4 = 1, y = 0) = \frac{3}{14}$$

$$P(x_4 = 0, y = 1) = \frac{3}{14} \qquad P(x_4 = 0, y = 0) = \frac{3}{14}$$

$$H(Y|x_4) = \sum_{x_4 \epsilon X} \sum_{y \epsilon Y} P(x_4, y) \log_2 \left(\frac{P(x_4)}{P(x_4, y)} \right)$$

$$= P(x_4 = 1, y = 1) \log_2 \left(\frac{P(x_4 = 1)}{P(x_4 = 1, y = 1)} \right) + P(x_4 = 0, y = 1) \log_2 \left(\frac{P(x_4 = 0)}{P(x_4 = 0, y = 1)} \right)$$

$$+ P(x_4 = 1, y = 0) \log_2 \left(\frac{P(x_4 = 1)}{P(x_4 = 1, y = 0)} \right) + P(x_4 = 0, y = 0) \log_2 \left(\frac{P(x_4 = 0)}{P(x_4 = 0, y = 0)} \right)$$

$$= \frac{5}{14} \log_2 \left(\frac{\frac{4}{7}}{\frac{5}{14}} \right) + \frac{3}{14} \log_2 \left(\frac{\frac{3}{7}}{\frac{3}{14}} \right) + \frac{3}{14} \log_2 \left(\frac{\frac{4}{7}}{\frac{3}{14}} \right) + \frac{3}{14} \log_2 \left(\frac{\frac{3}{7}}{\frac{3}{14}} \right)$$

$$= \frac{5}{14} \log_2 \left(\frac{8}{5} \right) + \frac{3}{14} \log_2 (2) + \frac{3}{14} \log_2 \left(\frac{8}{3} \right) + \frac{3}{14} \log_2 (2)$$

$$= \frac{5}{14} \log_2 \left(\frac{8}{5} \right) + \frac{3}{14} + \frac{3}{14} \log_2 \left(\frac{8}{3} \right) + \frac{3}{14}$$

$$= \frac{5}{14} \log_2 \left(\frac{8}{5} \right) + \frac{3}{14} \log_2 \left(\frac{8}{3} \right) + \frac{6}{14}$$

$$= 0.24216 + 0.30322 + 0.42857$$

$$\boxed{H(Y|x_4) = 0.97395}$$

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

$$I(x_1, Y) = H(Y) - H(Y|x_1) \qquad I(x_4, Y) = H(Y) - H(Y|x_4)$$

$$I(x_1, Y) = 0.98523 - 0.60458 \qquad I(x_4, Y) = 0.98523 - 0.97395$$

$$\boxed{I(x_1, Y) = 0.38065} \qquad \boxed{I(x_4, Y) = 0.01128}$$

Because a larger mutual information value means a larger reduction in uncertainty, we can say that x_1 is more informative.

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

$$Y = \{yes, no\} = \{1, 0\} \qquad x_3 = \{yes, no\} = \{1, 0\}$$

$$P(y = 1, x_3 = 1) = \frac{5}{14} \qquad P(y = 1, x_3 = 0) = \frac{3}{14}$$

$$P(y = 0, x_3 = 1) = \frac{2}{14} = \frac{1}{7} \qquad P(y = 0, x_3 = 0) = \frac{4}{14} = \frac{2}{7}$$

$$H(Y, x_3) = \sum_{y \epsilon Y} \sum_{x_3 \epsilon X} P(y, x_3) \log_2 \left(\frac{1}{P(y, x_3)} \right)$$

$$= -\sum_{y \epsilon Y} \sum_{x_3 \epsilon X} P(y, x_3) \log_2 (P(y, x_3))$$

$$= -P(y = 1, x_3 = 1) \log_2(P(y = 1, x_3 = 1)) - P(y = 0, x_3 = 1) \log_2(P(y = 0, x_3 = 1))$$

$$- P(y = 1, x_3 = 0) \log_2(P(y = 1, x_3 = 0)) - P(y = 0, x_3 = 0) \log_2(P(y = 0, x_3 = 0))$$

$$= -\frac{5}{14} \log_2 \left(\frac{5}{14} \right) - \frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{3}{14} \log_2 \left(\frac{3}{14} \right) - \frac{2}{7} \log_2 \left(\frac{2}{7} \right)$$

$$= 0.5305 + 0.40105 + 0.47622 + 0.51638$$

$$\boxed{H(Y, x_3) = 1.92417}$$

5.3 Entropy Proofs [7pts]

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

Since X and Y are independent, then $I(X,Y) = 0$. Therefore:

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

$$0 = H(X) - H(X|Y)$$

$$\boxed{H(X|Y) = H(X)}$$

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

$$H(X,Y) = H(X) + H(Y)$$

$$-\sum_x \sum_y P(x,y) \log_2 P(x,y) = -\sum_x P(x) \log_2 P(x) - \sum_y P(y) \log_2 P(y)$$

Since X and Y are independent, then: $P(X) = \sum_y P(x,y), \quad P(Y) = \sum_x P(y,x)$

$$-\sum_x \sum_y P(x,y) \log_2 P(x,y) = -\sum_x \sum_y P(x,y) \log_2 P(x) - \sum_y \sum_x P(y,x) \log_2 P(y)$$

Since:

$$\sum_x \sum_y P(x,y) = \sum_y \sum_x P(y,x)$$

$$-\sum_x \sum_y P(x,y) \log_2 P(x,y) = -\sum_x \sum_y P(x,y) \log_2 P(x)P(y)$$

Since X and Y are independent, then:

$$P(x)P(y) = P(x,y)$$

$$\boxed{-\sum_x \sum_y P(x,y) \log_2 P(x,y) = -\sum_x \sum_y P(x,y) \log_2 P(x,y)}$$

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

$$I(X,Y) = I(Y,X)$$

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

$$-\sum_{x} P(x)\log_2 P(x) + \sum_{x,y} P(x,y)\log_2 P(x|y) = -\sum_{y} P(y)\log_2 P(y) + \sum_{y,x} P(y,x)\log_2 P(y|x)$$

$$-\sum_{x,y} P(x,y)\log_2 P(x) + \sum_{x,y} P(x,y)\log_2 P(x|y) = -\sum_{y,x} P(y,x)\log_2 P(y) + \sum_{y,x} P(y,x)\log_2 P(y|x)$$

$$\sum_{x,y} P(x,y)\log_2\left(\frac{P(x|y)}{P(x)}\right) = \sum_{y,x} P(y,x)\log_2\left(\frac{P(y|x)}{P(y)}\right)$$

$$\boxed{\sum_{x,y} P(x,y)\log_2\left(\frac{P(x,y)}{P(x)P(y)}\right) = \sum_{y,x} P(y,x)\log_2\left(\frac{P(y,x)}{P(y)P(x)}\right)}$$

Because joint probability is commutative $P(x,y) = P(y,x)$

6 Bonus for All [10 pts]

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

From the definition of a Poisson distribution we know that:

$E[X] = Var[X] = \lambda$, which in this case is equal to 8.

$$
\begin{aligned}
E[(X+2)^2] &= E[X^2 + 4X + 4] \\
&= E[X^2] + 4E[X] + E[4] \\
&= Var[X] + (E[X])^2 + 4E[X] + E[4] \\
&= 8 + 64 + 32 + 4
\end{aligned}
$$

$$\boxed{E[(X+2)^2] = 108}$$

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

Here we have a geometric distribution with $p = \frac{1}{2}$. From the definition of a geometric series the variance is:

$$Var[Y] = \frac{p}{(1-p)^2}$$

$$Var[Y] = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2}$$

$$Var[Y] = \frac{\frac{1}{2}}{\frac{1}{4}}$$

$$\boxed{Var[Y] = \frac{4}{2} = 2}$$

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

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

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

$$P(X+Y \le 1) = \int_0^1 \int_0^{1-x} f_{x,y}(x,y)\, dy\, dx$$

$$= \int_0^1 \int_0^{1-x} (x+y)\, dy\, dx$$

$$= \int_0^1 (xy + \frac{y^2}{2})\Big|_0^{1-x} dx$$

$$= \int_0^1 (x(1-x) + \frac{(1-x)^2}{2})\, dx$$

$$= \int_0^1 (x - x^2 + \frac{1}{2} - x + \frac{x^2}{2})\, dx$$

$$= \int_0^1 (\frac{1}{2} - \frac{x^2}{2})\, dx$$

$$= \frac{x}{2} - \frac{x^3}{6}\Big|_0^1$$

$$= \frac{1}{2} - \frac{1}{6}$$

$$\boxed{P(X+Y \le 1) = \frac{1}{3}}$$