Summer 2021 CS4641/CS7641 A Homework 1

Dr. Mahdi Roozbahani
written by Payman Behnam

Deadline: June 9, Wednesday, 11:59 pm AOE

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

Instructions

- This assignment has no programming, only written questions.
- We will be using Gradescope this semester for submission and grading of assignments.
- Your write up must be submitted in PDF form, you can use any word processing package such as Latex, markdown, or any software you prefer. We will not accept handwritten work. (Links to tutorials - Undergrad access, Grad access)
- 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**. The calculation process means including the major steps to arrive at the final answer.

1 Linear Algebra [15pts + 8pts + 10pts]

1.1 Determinant and Inverse of Matrix [15pts]

Given a matrix M:

$$M = \begin{bmatrix} r & 6 & 0 \\ 2 & 6 & 1 \\ 1 & 1 & r \end{bmatrix}$$

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

Solution: The determinant of 2x2 Matrix is $det(\begin{bmatrix} a & b \\ c & d \end{bmatrix})$ ad - bc. So for the question The determinant is computed using the following formula.

$$(r) \times det(\begin{bmatrix} 6 & 1 \\ 1 & r \end{bmatrix}) - (6) \times det(\begin{bmatrix} 2 & 1 \\ 1 & r \end{bmatrix}) + (0) \times det(\begin{bmatrix} 2 & 6 \\ 1 & 11 \end{bmatrix})$$

$=$

$$6r^2 - 13r + 6$$

(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: If determinant is zero, the inverse will not exist. That means that

$$6r^2 - 13r + 6 = 0$$

So, the value for r would be $\frac{3}{2}$ and $\frac{2}{3}$. For both value of r, the matrix M is singular and the rank of the matrix will reduce by 1. So, it is 2 for both cases.

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

Solution: by using Cofactor method, we would have:

$$\begin{bmatrix} det(\begin{bmatrix} 6 & 1 \\ 1 & 4 \end{bmatrix}) & -det(\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}) & det(\begin{bmatrix} 2 & 6 \\ 1 & 1 \end{bmatrix}) \\ -det(\begin{bmatrix} 6 & 0 \\ 1 & 4 \end{bmatrix}) & det(\begin{bmatrix} 4 & 0 \\ 1 & 4 \end{bmatrix}) & -det(\begin{bmatrix} 4 & 6 \\ 1 & 1 \end{bmatrix}) \\ det(\begin{bmatrix} 6 & 0 \\ 6 & 1 \end{bmatrix}) & -det(\begin{bmatrix} 4 & 0 \\ 2 & 1 \end{bmatrix}) & det(\begin{bmatrix} 4 & 6 \\ 2 & 6 \end{bmatrix}) \end{bmatrix}$$

$=$

$$\begin{bmatrix} (6 \times 4 - 1 \times 1) & -(2 \times 4 - 1 \times 1) & (2 \times 1 - 6 \times 1) \\ -(6 \times 4 - 0 \times 1) & (4 \times 4 - 0 \times 1) & -(4 \times 1 - 6 \times 1) \\ (6 \times 1 - 0 \times 6) & -(4 \times 1 - 0 \times 2) & (4 \times 6 - 6 \times 2) \end{bmatrix}$$

=

$$\begin{bmatrix} 23 & -7 & -4 \\ -24 & 16 & 2 \\ 6 & -4 & 12 \end{bmatrix}$$

The transpose of the Matrix is

$$\begin{bmatrix} 23 & -24 & 6 \\ -7 & 16 & -4 \\ -4 & 2 & 12 \end{bmatrix}$$

So the inverse of matrix is: $\frac{1}{det(M)}M^T$ and det(M) based on the formula $6r^2 - 13r + 6$ when r=4 is 50. So we have:

$$\frac{1}{50)}\begin{bmatrix} 23 & -24 & 6 \\ -7 & 16 & -4 \\ -4 & 2 & 12 \end{bmatrix}$$

=

$$\begin{bmatrix} \frac{23}{50} & \frac{-12}{25} & \frac{3}{25} \\ \frac{-7}{50} & \frac{8}{25} & \frac{-2}{25} \\ \frac{-2}{25} & \frac{1}{25} & \frac{6}{25} \end{bmatrix}$$

$$MM^{-1} = \begin{bmatrix} 4 & 6 & 0 \\ 2 & 6 & 1 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} \frac{23}{50} & \frac{-12}{25} & \frac{3}{25} \\ \frac{-7}{50} & \frac{8}{25} & \frac{-2}{25} \\ \frac{-2}{25} & \frac{1}{25} & \frac{6}{25} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(d) Find the determinant of M^{-1} for $r = 4$. What is the relationship between the determinant of M and the determinant of M^{-1}? [3pts]

Solution: We follow the same procedure for $M^(-1)$ as well. The determinant is:

$$(\frac{23}{50}) * det(\begin{bmatrix} \frac{8}{25} & \frac{-2}{25} \\ \frac{1}{25} & \frac{6}{25} \end{bmatrix}) - \frac{12}{25} * det(\begin{bmatrix} \frac{-7}{50} & \frac{-2}{25} \\ \frac{-2}{25} & \frac{6}{25} \end{bmatrix}) + \frac{3}{25} * det(\begin{bmatrix} \frac{-7}{50} & \frac{8}{25} \\ \frac{-2}{25} & \frac{1}{25} \end{bmatrix})$$

=

$$\frac{1}{50}$$

$det(M) \times det(M^{-1}) = det(M) \times \frac{1}{det(M)} = 50 \times \frac{1}{50} = 1.$

Hence $det(M^{-1}) = \frac{1}{det(M)}$

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

Solution: We know a matrix is not invertible if and only if its determinant is 0. Since

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

, we have

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

So:

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

; then $(A - \lambda I) = 0$

If $(A - \lambda I)$ has an inverse, then we should have $(A - \lambda I) \times (A - \lambda I)^{-1} = I$. However, $(A - \lambda I) = 0$. So, the result of the multiplication $(A - \lambda I) \times (A - \lambda I)^{-1} = I$ would be zero any how.

This means that $(A - \lambda I)$ is not invertible (zero matrix does not have inverse). We know a matrix is not invertible if and only if its determinant is 0. So, $|A - \lambda I| = 0$ (singular matrices have zero determinant)

1.3 Eigenvalues and Eigenvectors [10pts]

Given a matrix A:

$$A = \begin{bmatrix} x & 9 \\ 4 & x \end{bmatrix}$$

(a) Calculate the eigenvalues of A as a function of x. (Calculation process required). [5 pts]

Solution: To find eigen pairs we have: $AX = \lambda X$ $|A - \lambda I| = \begin{vmatrix} x - \lambda & 9 \\ 4 & x - \lambda \end{vmatrix}$ $= (x - \lambda)^2 - 36$. If $|A - \lambda I| = 0$ we would have $\lambda = x + 6$ or $\lambda = x - 6$

(b) Find the normalized eigenvectors of matrix A (Calculation process required). [5 pts]

Solution:

$Av = \lambda v$ and $\lambda = x + 6$ or $\lambda = x - 6$ and

$$A = \begin{bmatrix} x & 9 \\ 4 & x \end{bmatrix}$$

So, $\begin{bmatrix} x & 9 \\ 4 & x \end{bmatrix} \begin{bmatrix} v1 \\ v2 \end{bmatrix}$ = (x+6) $\begin{bmatrix} v1 \\ v2 \end{bmatrix}$ or $\begin{bmatrix} x & 9 \\ 4 & x \end{bmatrix} \begin{bmatrix} v1 \\ v2 \end{bmatrix}$ = (x-6) $\begin{bmatrix} v1 \\ v2 \end{bmatrix}$

So, $v1 = \frac{3}{2}v2$ or $v1 = \frac{-3}{2}v2$. So, the normalized eigenvectors are:

if $\lambda = x + 6$ then the normalized eigenvalue is $\begin{bmatrix} \frac{3}{\sqrt{13}} \\ \frac{2}{\sqrt{13}} \end{bmatrix}$

if $\lambda = x - 6$ then the normalized eigenvalue is $\begin{bmatrix} \frac{-3}{\sqrt{13}} \\ \frac{2}{\sqrt{13}} \end{bmatrix}$

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

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 $Z = XY$.

(a) Show that Z also follows a Normal (Gaussian) distribution. Calculate the Expectation and Variance of Z. [9pts]

Solution We know Z is a function of X and Y. For a fixed z_i value we have: $z_i = x_i * y_i$ and $z_i = -x_i * -y_i$. So:

$$P(Z) = P(X)P(Y|X) + P(-X)P(-Y|-X)$$

Since X and Y are independent: $P(Z) = P(X)P(Y) + P(-X)P(-Y) = 0.5P(Y) + 0.5P(-Y)$

P(Y) and P(-Y) have normal distributions. We know the sum of two normal distributions is also a normal distributions, hence Z has a normal distribution.

To calculate the mean of X we have:

$$E(X) = \sum_{n=1} xp(x) = (c \times 0.5) - (c \times 0.5) = 0$$

To calculate variance of X we have:

$$var(X) = E[(X - \mu)^2] = E[(X - 0)^2] = (0.5 \times c^2) + (0.5 \times (-c)^2) = c^2$$

To calculate the mean of Z we have:

$$E[Z] = \int_y \sum_x p(x, y)xydy$$

since X and Y are independent: $E[Z] = \int_y p(y)ydy \sum_x p(x) = E[Y] \times E[X] = 0 \times 0$

To calculate variance of Z we have:

$$var(Z) = E[Z^2] - E[Z]^2 = E[X^2Y^2] - 0 = \int_y \sum_x p(x, y)x^2y^2dy$$

since X and Y are independent: $var(Z) = \int_y p(y)y^2dy \sum_x p(x)x^2 = var(Y) \times var(X) = 1 \times c^2 = c^2$

(b) How should we choose c such that Y and Z are uncorrelated? [5pts]

Solution:

$$cov(Y, Z) = E[YZ] - E[Y]E[Z]$$

since $Z = XY$:

$$cov(Y, Z) = E[Y^2 X] - 0 = \int_y \sum_x p(x, y) x y^2 dy$$

since X and Y are independent:

$$cov(Y, Z) = \int_y p(y) y^2 dy \sum_x p(x) x = var(Y) \times E[X] = 0$$

This means that Z and Y are always uncorrelated. So, the value of c doesn't matter as long as its value is real and finite.

(c) Let X and Y be independent random variables with $var(X) = 3$ and $var(Y) = 4$. We do not know $E[X]$ or $E[Y]$. Let $Z = 2X + Y$. What is the correlation coefficient $\rho(X, Z) = \frac{cov(X,Z)}{\sqrt{var(X)var(Z)}}$? **Note: 2(c) has no connection with 2(a) and 2(b)** [4pts]

Solution: Since X and Y are independent we have:

$$var(Z) = E[Z^2] - E[Z]^2 = E[(2X + Y)^2] - E[2X + Y]^2$$

$=$

$$4(E[X^2] - E[X]^2) + E[Y^2] - E[Y]^2 = 4var(X) + var(Y) = 4 \times 3 + 4 = 16$$

Also we know

$$Cov(X, Z) = E[XZ] - E[X]E[Z] = E[X(2X + Y)] - E[X]E[2X + Y]$$

$=$

$$E[X^2] + E[X]E[Y] - 2E[X]^2 - E[X]E[Y]$$

$=$

$$2(E[X^2] - E[X]^2) = 2Var(X) = 2 \times 3 = 6$$

So:

$$\rho(X, Z) = \frac{cov(X, Z)}{\sqrt{var(X)var(Z)}} = \frac{6}{\sqrt{3 \times 16}} = 0.86$$

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:

$$\begin{aligned} \max_{x,y} \quad & f(x,y) = 2x + 5y^2 \\ \text{s.t.} \quad & g_1(x,y) = x^2 + y^2 \leq 4 \\ & g_2(x,y) = y \leq 1 \end{aligned}$$

(a) Specify the Lagrange function [2 pts]

Solution

The Lagrangian is $L(x, y, \lambda_1, \lambda_2) = f(x) + \lambda_1(g_1(x)) + \lambda_2(g_2(x)) = 2x + 5y^2 + \lambda_1(x^2 + y^2 - 4) + \lambda_2(y - 1)$

(b) List the KKT conditions [2 pts]

Solution

Stationary:

$$\begin{bmatrix} 2 \\ 10y \end{bmatrix} + \lambda_1 \begin{bmatrix} 2x \\ 2y \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

So we have:

$$\begin{bmatrix} 2 + 2\lambda_1 x \\ 10y + 2\lambda_1 y + \lambda_2 \end{bmatrix} = 0$$

Primal feasibility:

$$x^2 + y^2 - 4 \leq 0$$

$$y - 1 \leq 0$$

Dual feasibility:

$$\lambda_1, \lambda_2 \geq 0$$

Complementary slackness:

$$\lambda_1(x^2 + y^2 - 4) = 0$$

and

$$\lambda_2(y - 1) = 0$$

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

Solution

1) Only nonlinear constraint is active:

$x^2 + y^2 - 4 = 0$ and $y - 1 \leq 0$.

$\lambda_2 = 0$ (complementary slackness) Further more, from complementary slackness,

$$\begin{bmatrix} 2 \\ 10y \end{bmatrix} + \lambda_1 \begin{bmatrix} 2x \\ 2y \end{bmatrix} = 0$$

So we have:

$$\begin{bmatrix} 2 + 2\lambda_1 x \\ 10y + 2\lambda_1 y \end{bmatrix} = 0$$

$x = -\frac{1}{\lambda_1}$ and $y(\lambda_1 + 5) = 0$ So, $y = 0$ or $\lambda_1 = -5$. The latter is not possible because Lagrange multipliers cannot be negative. Hence, y = 0 is acceptable. From nonlinear constraint, $x = +2$ or $x = -2$. Since $\lambda_1 \geq 0$, $x_1 = -2$. so we have $(-2, 0)$ and $f(-2, 0) = 2 \times (-2) + 0 = -4$

2) Only linear constraint is active:

$x^2 + y^2 - 4 = 0$ and $y = 1$.

$\lambda_1 = 0$ (complementary slackness) Further, from complementary slackness,

$$\begin{bmatrix} 2 \\ 10y \end{bmatrix} + \lambda_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

So we have:

$$\begin{bmatrix} 2 \\ 10y + \lambda_2 \end{bmatrix} = 0$$

So, no solution exist!

3) Both constraints are active:

$x^2 + y^2 - 4 = 0$ and $y - 1 = 0$. So, $y = 1$ and $x = \sqrt{(3)}$ and $x = -\sqrt{(3)}$ so we have $(\sqrt{(3)}, 1)$, $(-\sqrt{(3)}, 1)$.

So: $f(\sqrt{(3)}, 1) = 5 + 2\sqrt{3}$

$f(-\sqrt{(3)}, 1) = 5 - 2\sqrt{3}$

The function is minimized at $f(-\sqrt{(3)}, 1)$

4) Both constraints are inactive: Due to complementary slackness $\lambda_1 = 0$ and $\lambda_2 = 0$

$$\begin{bmatrix} 2 \\ 10y \end{bmatrix} = 0$$

So, no solution exist.

(d) List all candidate points [4 pts]

Solution:

All candidate points are $(\sqrt(3), 1)$, $(-\sqrt(3), 1)$, $(-2, 0)$

(e) Find the candidate point for which f(x,y) has the maximum value. Check if L(x,y) if concave or convex at this point. [2 pts]

Solution:

$$f(\sqrt{3}, 1) = 5 + 2\sqrt{3}$$

$$f(-\sqrt{3}, 1) = 5 - 2\sqrt{3}$$

$$f(-2, 0) = -4$$

The function gets its max in $(\sqrt{3}, 1)$ and min in (-2,0)

To decide if L(x,y) is convex or concave, we need to computer the second derivative of L(x,y). If positive it is convex and if negative is is concave. alternatively we can compute Hessian matrix. At (-2,0) the function is convex as hessian is positive.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

For computing Hessian I used Hessian calculator: https://www.wolframalpha.com/widgets/view.jsp?id=3d7a13b30d8ebcdf80063f3d5633f18c

4 Maximum Likelihood [10 + 15 + 15 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	Tail

(a) What is the likelihood of the result given θ? [3pts]

Solution The results for A are 5 Tails and 2 Heads and for B are 2 Heads and 1 Tail. So:

$$L(\theta) = \theta^2 \times (1-\theta)^5 \times (2\theta)^2 \times (1-2\theta) = (1-\theta)^5 \times (1-2\theta) \times 4\theta^4$$

(b) What is/are the maximum likelihood estimation(s) for θ? [7pts]

$$L(\theta) = (1-\theta)^5 \times (1-2\theta) \times 4\theta^4$$

$$\log L(\theta) = 5\log(1-\theta) + \log(1-2\theta) + \log 4 + 4\log\theta$$

$$\frac{\sigma \log L(\theta)}{\sigma\theta} = 0$$

So:

$$\frac{4}{\theta} - \frac{5}{1-\theta} - \frac{2}{1-2\theta} = 0$$

Accordingly:

$$\frac{20\theta^2 - 19\theta + 4}{(\theta-1)(\theta)(2\theta-1)} = 0$$

S0 we have: $\theta = \frac{19}{40} + \frac{\sqrt{41}}{40}$ or $\theta = \frac{19}{40} - \frac{\sqrt{41}}{40}$.

$\theta = \frac{19}{40} + \frac{\sqrt{41}}{40}$ makes the probability $2\theta \geq 0$, which is impossible!

4.2 Normal distribution [15 pts]

Suppose we are given some points (x, y). Suppose we know that there is a linear relationship between x and y, that is, $y = mx + c$ for some m, c. However, as is often the case with real-world data, we can only observe corrupted outputs $y' = y + z$, where z is noise. Here y represents the real values and y' are the observed, noisy values.

We assume that the noise is Gaussian with mean 0 and variance σ^2. So,

$$z \sim Normal\left(0, \sigma^2\right)$$

1. Assume there is just one sample (x_1, y'_1). Given x_1, m and c, find the likelihood of this sample. [5 pts]

   Solution:

   $y_1 = mx_1 + c$, and $y' = y + z$. So: $z = y'_1 - mx_1 - c$.

   Accordingly, $P(y'|m, c, x_1) = P(y' - y_1|m, c, x1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y'_1 - mx_1 - c)^2}{2\sigma^2}}$

2. Now, assume we are given n such values (x_i, y'_i). Assume $c = 0$. What is the likelihood of this data given m? [5 pts]

   since $c = 0$ we have $P(y'_i|m, x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y'_i - mx_i)^2}{2\sigma^2}}$. Since we have n sample: $P(y'|m, x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y' - mx)^2}{2\sigma^2}} = [\frac{1}{\sqrt{2\pi\sigma^2}}]^n e^{-\sum_{i=1}^{n} \frac{(y' - mx)^2}{2\sigma^2}}$

3. Find the maximum likelihood estimator for m. [5 pts]
   Hint: To maximize the likelihood, what should be the derivative of log likelihood?

   Solution:

   $$\log AB = \log A + \log B$$

   The base of log is e

   $$\log P(y'|m, x) = \log([\frac{1}{\sqrt{2\pi\sigma^2}}]^n e^{-\sum_{i=1}^{n} \frac{(y' - mx)^2}{2\sigma^2}}) =$$

   $$-n \log(\sqrt{2\pi\sigma^2} - \sum_{i=1}^{n} \frac{(y' - mx)^2}{2\sigma^2}$$

   We take derivative and make it zero.

   $$\frac{\sigma \log P(y'|m, x)}{\sigma m} = 0$$

   Hence:

   $$-2 \sum_{i=1}^{n} \frac{(y'_i - mx_i)(-x_i)}{2\sigma^2} = 0$$

   . MLE value is $\frac{\sum_{i=1}^{n} y' x}{\sum_{i=1}^{n} x_i^2}$

4.3 Exponential Distribution: Bonus for undergrads [15 pts]

There are n independent random variables $X_1, X_2, ..., X_n$. The C.D.F of the variables is

$$P(X_i \leq x|\lambda_i) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda_i x}, & 0 \leq x \end{cases}$$

where $\lambda_i \geq 0$.

1. Write down the P.D.F of above independent random variables. [2pts]

   Solution: By the theorem of calculus, the pdf is obtained by differentiating the cdf: $f(x) = \frac{dF(x)}{dx}$. So:

   $$f(x) = \begin{cases} 0 & x < 0 \\ \frac{d(1-e^{-\lambda_i x})}{dx} = \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

2. Calculate the expected value of distribution for a given i. (No need to show the calculations) [3 pts]

   Solution:

   $E[Xi] = \int_{-\infty}^{\infty} \lambda e^{-\lambda x}\, dx = \int_{0}^{\infty} \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}$

3. Suppose we are given observations $(x_1, x_2, ..., x_n)$. Assume n is even and all values are positive. Suppose

   $$\lambda_i = \begin{cases} \lambda, & 0 \leq i \leq n/2 \\ 2\lambda, & n/2 + 1 \leq i \leq n \end{cases}$$

   Find Maximum Likelihood Estimator for λ, given these values. [6pts]

   Hint: To maximize the likelihood, what should be the derivative of log likelihood? [6pts]

   Solution:

   $$P(Xi = xi) = \begin{cases} \lambda e^{-\lambda x_i}, & 0 \leq i \leq n/2 \\ 2\lambda e^{-2\lambda x_i}, & n/2 + 1 \leq i \leq n \end{cases}$$

   $L(x) = \prod_{i=0}^{n/2} \lambda e^{-\lambda x} \prod_{i=n/2+1}^{n} 2\lambda e^{-2\lambda x} = 2^{n/2} \lambda^n e^{-\lambda[\sum_{i=0}^{n/2} x_i + \sum_{i=n/2+1}^{n} 2x_i]}$

   $\log(2^{\frac{n}{2}}) + n \log \lambda - \lambda[\sum_{i=0}^{n/2} x_i + \sum_{i=n/2+1}^{n} 2x_i]$

   By derivation we have:

   $$\frac{\sigma \log L(x)}{\sigma \lambda} = 0$$

   $0 + \frac{n}{\lambda} - [\sum_{i=0}^{n/2} x_i + \sum_{i=n/2+1}^{n} 2x_i] = 0$

   So we have: $\lambda = \frac{n}{[\sum_{i=0}^{n/2} x_i + \sum_{i=n/2+1}^{n} 2x_i]}$

4. Assume $\lambda_1 = \lambda$. What is the MLE for λ for $n = 1$ variable? Can you find a relation between this value and the expected value of the variable? [4 pts]

   Solution:

   In this case, we have only one variable. So we have

   $$L(x_0) = \lambda e^{(-\lambda x_0)}$$

   By derivation we have:

   $$\frac{\sigma \log L(x)}{\sigma \lambda} = 0$$

   $\frac{1}{\lambda} - x_0 = 0$ so: $\lambda = \frac{1}{x_0}$

   On the other hand $E[X1] = \int_0^x \lambda x e^{-\lambda x}\, dx \ = \frac{1}{\lambda} = x_0$

   So, the variable and its expected value should be equal.

5 Information Theory [35 pts]

5.1 Marginal Distribution [5pts]

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.

$X \mid Y$	0	1
0	$\frac{1}{5}$	$\frac{1}{5}$
1	$\frac{2}{5}$	$\frac{1}{5}$

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

Solution:

$$P(X=0) = P(X=0|Y=0) + P(X=0|Y=1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

$$P(X=1) = P(X=1|Y=0) + P(X=1|Y=1) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

$$P(Y=0) = P(Y=0|X=0) + P(Y=0|X=1) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$$

$$P(Y=1) = P(Y=1|X=0) + P(Y=1|X=1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

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

$X \mid Y$	0	1	
0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
1	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{3}{5}$
	$\frac{3}{5}$	$\frac{2}{5}$	1

Solution:

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

$H(Y) = P(Y=0)\log\frac{1}{P(Y=0)} + P(Y=1)\log\frac{1}{P(Y=1)} = \frac{3}{5} \times \log\frac{5}{3} + \frac{2}{5} \times \log\frac{5}{2} = 0.97$

$H(Y|X=0) = 2 \times \frac{1}{2}\log(2) = 1$

$H(Y|X=1) = \frac{2}{3} \times \log(\frac{3}{2}) + \frac{1}{3} \times \log(3) = 0.92$

$H(Y|X) = P(X=0)H(Y|X=0) + P(X=1)H(Y|X=1) = \frac{2}{5} \times 1 + \frac{3}{5} \times 0.92 = 0.95$

So: $I(X,Y) = 0.02$

5.2 Mutual Information and Entropy [20pts]

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 does not quarantine) is represented as Y.

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

Solution:

In case of no: $P(Y) = \frac{6}{14}$. In case of yes: $P(Y) = \frac{8}{14}$.

$H(Y) = H(8/14, 6/14) = 0.98$

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

Solution:

$$P(Y = yes|x1 = young) = \frac{1}{5}$$
$$P(Y = yes|x1 = senior) = \frac{3}{5}$$
$$P(Y = yes|x1 = middleaged) = 1$$

$$H(Y|x1)$$

$$H(Y|x1 = young) = H(\frac{1}{5}, \frac{4}{5}) = 0.72$$

$$H(Y|x1 = senior) = H(\frac{3}{5}, \frac{2}{5}) = 0.97$$
$$H(Y|x1 = middleaged) = H(1, 0) = 0$$

$H(Y|x1) = P(x1 = young) \times H(Y|x1 = young) + P(x1 = senior) \times H(Y|x1 = senior) + P(x1 = middleage$

$=$

$$\frac{5}{14} \times 0.72 + \frac{5}{14} \times 0.97 + 0 = 0.60$$

$$H(Y|x4)$$

$H(Y|x4) = P(x4 = yes) \times H(Y|x4 = yes) + P(x4 = no) \times H(Y|x4 = no)$

$=$

$$\frac{8}{14} \times 0.9544 + \frac{6}{14} \times 1 = 0.97$$

(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]

$$I(x1, Y) = H(Y) - H(Y|x1) = 0.98 - 0.60 = 0.38$$
$$I(x4, Y) = H(Y) - H(Y|x4) = 0.98 - 0.97 = 0.01$$

Since $I(x_1, Y) > I(x_4, Y)$, x1 provides more information.

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

$H(Y, x3) = P(Y = yes, x3 = yes) \times \log(\frac{1}{P(Y=yes,x3=yes)}) + P(Y = no, x3 = no) \times \log(\frac{1}{P(Y=no,x3=no)}) + P(Y = no, x3 = yes) \times \log(\frac{1}{P(Y=no,x3=yes}))+ P(Y = yes, x3 = no) \times \log(\frac{1}{P(Y=yes,x3=no)}) = 1.92$

5.3 Entropy Proofs [10 pts]

(a) Prove that $I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

Solution1 First Proof:

$H(X) - H(X|Y) = -\sum_x pX(x)\log(pX(x)) + \sum_y pY(y)\sum_x[pX|Y(x,y)\log(pX|Y(x,y))] =$
$\sum_y\sum_x pX,Y(x,y)\log(pX|Y(x,y))] - \sum_y\sum_x pX,Y(x,y)\log(pX(x)) = \sum_y\sum_x pX,Y(x,y)\log[\frac{pX,Y(x,y)}{(pY(y)pX(x}$
$\sum_y\sum_x pX,Y(x,y)\log[\frac{pY|X(x,y)}{(pY(y)}] = -\sum_x\sum_y pX,Y(x,y)\log(pY(y)) + \sum_x\sum_y pX,Y(x,y)\log(pY|X(x,y)$

The last statement is exactly the one used for $H(X) - H(X|Y)$ with replacing x with y and vice versa. So that means the last one is equal to $H(Y) - H(Y|X)$

Solution2

Second Proof: We know I(X,Y) and I(Y,X) are equal. So, by changing X to Y and Y to X we have:

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

(b) Prove that $H(X,Y) = H(X|Y) + H(Y|X) + I(X,Y)$.
Hint: You can use the formula for relation between joint and conditional entropy to deduce one part from the other.

Solution

$$1) I(X,Y) = H(Y) - H(Y|X)$$

$$2) I(X,Y) = H(X) - H(X|Y)$$

By adding (1), (2) we have: $3) 2I(X,Y) = H(X) - H(X|Y) + H(Y) - H(Y|X)$

Also we know :

$$4) H(X,Y) = H(X) + H(Y|X)$$

$$5) H(X,Y) = H(Y) + H(X|Y)$$

By adding (4), (5) we have: $6) 2H(X,Y) = H(X) + H(X|Y) + H(Y) + H(Y|X)$

By subtracting (3) from (6) we have:

$$2H(X,Y) - 2I(X,Y) = 2H(X|Y) + 2H(Y|X)$$

So:

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

6 Bonus for All [20 pts]

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

Solution For Poisson distribution, Var(X) = E[X]. With that we have: $E[(2X-3)^2] = E[4X^2 - 12X + 9] = 4E[X^2] - 12E[X] + E[9] = 4(E[X^2] - E[X]^2) + 4E[X]^2 - 12E[X] + E[9] = 4Var(X) + 4 \times 5^2 + 9 - 60 = 69$

(b) Suppose that X and Y have joint pdf given by [5 pts]

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

What are the marginal probability density functions for X and Y?

Solution

It is obvious for $x < 0, f_X(x) = 0$, otherwise when $0 \leq x \leq 1$ we have: $f_X(x) = \int_{y=0}^{\infty} f_{X,Y}(x,y)\, dy \ = \int_{y=0}^{\infty} 2e^{-2y}\, dy \ = 1$

Also, It is obvious for $y < 0, f_Y(y) = 0$, otherwise when $y \leq 0$ we have:

$f_Y(y) = \int_{x=0}^{\infty} f_{X,Y}(x,y)\, dx \ = \int_{y=0}^{\infty} 2e^{-2y}\, dx \ = 2e^{-2y}$

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

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

What is the probability P(X+Y $\leq$ 1)? [5 pts]

Solution

$P(X+Y \leq 1) = \int_{x=0}^{1}\int_{y=0}^{1-x} f_{X,Y}(x,y)\, dxdy \ = \int_{x=0}^{1}\int_{y=0}^{1-x} (x+y)\, dxdy \ = \int_{x=0}^{1}(\frac{1+x^2-2x}{2} + x - x^2)\, dxdy \ = \frac{1}{3}$

(d) You are observing an intersection that wild llamas frequently pass through. You are stationary. Each time a llama arrives at this intersection you write down the time of arrival – you notice these arrival times follow a Poisson distribution. Lamas arrive at this intersection at a rate of 900 lamas per hour. Let us define a gap in llama arrivals as the time difference between two subsequent (one-after-the-other) llamas arriving at the intersection. Find the probability that a gap in llama arrivals is greater than 8 seconds. [5 pts]

Solution

Lamas arrival rate at the intersection $\frac{900}{1h} = \frac{900}{3600s} = \frac{1}{4} = 0.25$ Accordingly, the pdf of the Lamas arrival rate is $0.25e^{-0.25}$

$P(X > 8) = \int_{8}^{\infty} 0.25e^{-0.25}\, dt \ = 0.13$