CS7641 A Homework 1

Sagarika Srishti: 903511279

September 11, 2020

1 Linear Algebra

1.1 Determinant and Inverse of Matrix

Given matrix M:

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

(a) Determinant of M.

We know that the determinant of a 3x3 matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is given by:

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Substituting the values for matrix M, we get:

$$\begin{aligned} |M| &= r \times 3 \times 3 + 6 \times r \times 4 + 0 \times 2 \times 7 - r \times r \times 7 - 6 \times 2 \times 3 - 0 \times 3 \times 4 \\ &= 9r + 24r - 7r^2 - 36 \\ &= -7r^2 + 33r - 36 \end{aligned}$$

(b) We know that a matrix M is singular or non-invertible if its determinant is 0. Thus, equating the value of determinant of M to zero, we get:

$$|M| = 0$$
$$-7r^2 + 33r - 36 = 0$$
$$7r^2 - 33r + 36 = 0$$
$$7r^2 - 21r - 12r + 36 = 0$$
$$7r(r-3) - 12(r-3) = 0$$
$$(7r-12)(r-3) = 0$$

Thus, M^{-1} does not exist for $r = 12/7$ and $r = 3$. For these values of r, the matrix M is not full-rank, as a matrix has to be full-rank for its inverse to exist. Thus, $rank(A) < 2$. This also means that M is singular for these values of r.

(c) Calculating M^{-1}:

For $r = 4$, matrix M becomes:

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

Determinant of M upon substituting the value or r:

$$\begin{aligned} |M| &= -7r^2 + 33r - 36 \\ &= -7 \times 16 + 33 \times 4 - 36 \\ &= -16 \end{aligned}$$

Now, we will find the classical adjoint of M.

$$((adj(M))_{ij} = (-1)^{i+j}|M_{\backslash j, \backslash i}|$$

$$adj(M) = \begin{bmatrix} \det\begin{vmatrix} 3 & 7 \\ 4 & 3 \end{vmatrix} & -\det\begin{vmatrix} 6 & 7 \\ 0 & 3 \end{vmatrix} & \det\begin{vmatrix} 6 & 3 \\ 0 & 4 \end{vmatrix} \\ -\det\begin{vmatrix} 2 & 4 \\ 4 & 3 \end{vmatrix} & \det\begin{vmatrix} 4 & 4 \\ 0 & 3 \end{vmatrix} & -\det\begin{vmatrix} 4 & 2 \\ 0 & 4 \end{vmatrix} \\ \det\begin{vmatrix} 2 & 4 \\ 3 & 7 \end{vmatrix} & -\det\begin{vmatrix} 4 & 4 \\ 6 & 7 \end{vmatrix} & \det\begin{vmatrix} 4 & 2 \\ 6 & 3 \end{vmatrix} \end{bmatrix}$$

$$adj(M) = \begin{bmatrix} (9-28) & -(18-0) & (24-0) \\ -(6-16) & (12-0) & -(16-0) \\ (14-12) & -(28-24) & (12-12) \end{bmatrix}$$

$$adj(M) = \begin{bmatrix} -19 & -18 & 24 \\ 10 & 12 & -16 \\ 2 & -4 & 0 \end{bmatrix}$$

Now, we know that:

$$M^-1 = adj(M)/|M|$$

Substituting the values of $adj(M)$ and $|M|$, we get:

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

To check if our inverse is correct, we will verify if $MM^{-1} = I$. For simplicity of calculation, we will divide by the value of $|M|$ at the end, so that we do not have to deal with fractions.

$$\begin{bmatrix} 4 & 6 & 0 \\ 2 & 3 & 4 \\ 4 & 7 & 3 \end{bmatrix} \times \begin{bmatrix} -19 & -18 & 24 \\ 10 & 12 & -16 \\ 2 & -4 & 0 \end{bmatrix} /16$$

$$= \begin{bmatrix} (4 \times 19 - 10 \times 6) & (4 \times 18 - 12 \times 6) & (-4 \times 24 + 6 \times 16) \\ (2 \times 19 - 3 \times 10 - 4 \times 2) & (2 \times 18 - 3 \times 12 - 4 \times 4) & (-2 \times 24 + 3 \times 16) \\ (4 \times 19 - 7 \times 10 - 3 \times 2) & (4 \times 18 - 7 \times 12 + 3 \times 4) & (-4 \times 24 + 7 \times 16) \end{bmatrix} /16$$

$$= \begin{bmatrix} (76 - 60) & (72 - 72) & (-96 + 96) \\ (38 - 30 - 8) & (36 - 36 + 16) & (-48 + 48) \\ (76 - 70 - 6) & (72 - 84 + 12) & (-96 + 112) \end{bmatrix} /16$$

$$= \begin{bmatrix} 16 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 16 \end{bmatrix} /16 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Hence, the inverse is correct.

(d) Determinant of M^{-1} for $r = 4$:
For $r = 4$, we know that

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

We know that the determinant of a 3x3 matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is given by:

$$|A| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Substituting the values for matrix M^{-1}, we get:

$$\begin{aligned} |M^{-1}| &= 19/16 \times 3/4 \times 0 - 9/8 \times 1 \times 1/8 + 3/2 \times 5/8 \times 1/4 - 19/16 \times 1 \times 1/4 + 9/8 \times 5/8 \times 0 + 19/16 \\ &= -9/64 + 15/64 - 19/64 + 9/6436 \\ &= -4/64 \\ &= -1/16 \end{aligned}$$

This could also have been proved by the property of determinants that:

$$|A^{-1}| = 1/|A|$$

1.2 Characteristic Equation

Given:

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

where x is a non-zero eigenvector and λ is eigenvalue of A.
My understanding of this proof follows from the Linear Algebra Review and Reference document that was a class reading (Linear Algebra Overview).

We are given:

$$Ax = \lambda x$$

This can be written as:

$$Ax - \lambda x = 0$$

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

Since λ is a scalar, it is multiplied by identity matrix I so that its dimensions are the same as matrix A. The above equation has a non-zero solution to x if and only if $(A - \lambda I)$ has linearly dependent columns. When a matrix has linearly dependent columns, it means that it is non-invertible, or singular. We know that a matrix is singular when its determinant is zero. Thus, the above equation has a non-zero solution to x if and only if $|(A - \lambda I)| = 0$.

This can also be proved by contradiction. If $(A - \lambda I)$ was invertible, we could've said:

$$(A - \lambda I)^{-1}(A - \lambda I)x = 0$$

$$Ix = 0$$

$$x = 0$$

This is a contradiction since the first equation states that $x \neq 0$. Therefore, $(A - \lambda I)$ is not invertible, which means that its determinant $|(A - \lambda I)|$ is 0.

Hence, proved!

1.3 Eigenvalues and Eigenvectors

Given matrix A:

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

(a) Eigenvalues of A as a function of x:
Let λ be the eigenvalues of A. We know that:

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

Substituting the value of matrix A, we get:

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

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

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

$$\lambda^2 + x^2 - 2x\lambda - 3 = 0$$

Using the equation for finding roots of a quadratic equation, we get:

- $$\begin{aligned} \lambda &= (2x + \sqrt{4x^2 - 4x^2 + 12})/2 \\ &= x + \sqrt{3} \end{aligned}$$

- $$\begin{aligned} \lambda &= (2x - \sqrt{4x^2 - 4x^2 + 12})/2 \\ &= x - \sqrt{3} \end{aligned}$$

Hence, the eigenvalues of A are $\lambda = x + \sqrt{3}$ and $\lambda = x - \sqrt{3}$.

(b) Normalized eigenvectors of matrix A:
To find the eigenvectors of A, we substitute the values of λ we found in the previous question in the equation

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

For $\lambda_1 = x + \sqrt{3}$, we get:

$$A - \lambda I = \begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} - \begin{bmatrix} x + \sqrt{3} & 0 \\ 0 & x + \sqrt{3} \end{bmatrix} = 0$$

$$\begin{bmatrix} -\sqrt{3} & 3 \\ 1 & \sqrt{3} \end{bmatrix} \begin{bmatrix} p_1 \\ q_1 \end{bmatrix}$$

This gives us:

$$p_1 = \sqrt{3} q_1$$

For normalized vectors,

$$p_1^2 + q_1^2 = 1$$

This means that

$$q1 = \pm 1/2$$

and

$$p1 = \pm\sqrt{3}/2$$

Similarly, for $\lambda_2 = x - \sqrt{3}$, we get:

$$A - \lambda I = \begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} - \begin{bmatrix} x - \sqrt{3} & 0 \\ 0 & x - \sqrt{3} \end{bmatrix} = 0$$

$$\begin{bmatrix} \sqrt{3} & 3 \\ 1 & -\sqrt{3} \end{bmatrix} \begin{bmatrix} p_2 \\ q_2 \end{bmatrix}$$

This gives us:

$$p_2 = -\sqrt{3} q_2$$

For normalized vectors,

$$p_2^2 + q_2^2 = 1$$

This means that

$$q2 = \pm 1/2$$

and

$$p2 = \mp\sqrt{3}/2$$

Hence, the normalized eigenvectors for matrix A are:
$x_1 = \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix}$ and $x_2 = \begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \end{bmatrix}$

2 Expectation, Co-variance and Independence

Given, X, Y and Z are three different random variables. X obeys a Bernouli 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. Y obeys a standard Normal (Gaussian) distribution, which can be written as $Y \sim N(0, 1)$. X and Y are independent. Meanwhile, $Z = XY$.

(a) Given,

$$Z = XY$$

Then,

$$\begin{aligned} P(Z < z) &= P(XY < z) \\ &= P(Y < z|X = c)P(X = c) + P(-Y < z|X = -c)P(X = -c) \end{aligned}$$

Since X and Y are independent, and the values of probabilities of X at these values is given, we get:

$$\begin{aligned} P(Z < z) &= P(Y < z|X = c)P(X = c) + P(-Y < z|X = -c)P(X = -c) \\ &= 0.5 \times P(Y < z) + 0.5 \times P(Y > -z) \\ &= 0.5(P(Y < z) + P(Y > -z) \end{aligned}$$

Since Y has normal distribution with mean 0, we know that it is symmetric about the y-axis. Hence,

$$\begin{aligned} P(Z < z) &= 0.5(P(cY < z) + P(cY > -z) \\ &= 0.5(2 \times P(cY < z)) \\ &= P(cY < z) \\ &= cY \end{aligned}$$

Hence, Z also follows a Normal distribution.

- We know that the expectation of product of two independent random variables is calculated as:

$$E[XY] = E[X] \times E[Y]$$

Calculating $E[X]$,

$$\begin{aligned} E[X] &= 0.5 \times c + 0.5 \times (-c) \\ &= 0.5c - 0.5c \\ &= 0 \end{aligned}$$

Hence,

$$\begin{aligned}E[Z] &= E[XY] \\ &= E[X]E[Y] \\ &= 0\end{aligned}$$

- We know that the variance of product of two independent random variables is calculated as:

  $$Var(XY) = Var(X)Var(Y) + Var(X)E^2[Y] + Var(Y)E^2[X]$$

  Now, we know that $E[X] = 0$ (from previous part), $E[Y] = 0$ (given) and $Var(Y) = 1$ (given).
  Calculating the other terms:

  $$\begin{aligned}Var(X) &= E[(X - E[X])^2] \\ &= E[X^2] \\ &= \sum_x x^2 \times p(x) \\ &= 0.5c^2 + 0.5c^2 \\ &= c^2\end{aligned}$$

  Substituting these values, we get:

  $$\begin{aligned}Var(XY) &= Var(X)Var(Y) + Var(X)E^2[Y] + Var(Y)E^2[X] \\ &= c^2 \times 1 + c^2 \times 0 + 1 \times 0 \\ &= c^2\end{aligned}$$

  Hence, $Var(Z) = c^2$.

(b) Choosing c such that Y and Z are uncorrelated(which means $Cov(Y, Z) = 0$):
We have,

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

We know that $Z = XY$ and X and Y are independent. Making use of these facts, we get,

$$\begin{aligned}
Cov(Y, Z) &= E[YZ] - E[Y]E[Z] \\
&= E[Y.XY] - E[Y]E[XY] \\
&= E[Y^2]E[X] - E[Y]^2E[X] \\
&= 0 - 0 \\
&= 0
\end{aligned}$$

We see that $Cov(Y, Z)$ is still 0. Thus, Y and Z are uncorrelated and this does not depend on values of c.

(c) Finding whether Y and Z are independent:
Since Y is a Normal variable with mean 0, we know that the probability $P(Y \in (-1, 0))$ can be calculated as the area under the curve between -1 and 0, which will be > 0.
Similarly, since Z is also a Normal variable, as proved in part a), $P(Z \in (2c, 3c))$ is also > 0.
However, since $y \notin (-1, 0)$, P(Y)=0. Thus,

$$\begin{aligned} P(Y, Z) &= P(Z|Y)P(Y) \\ &= 0 \end{aligned}$$

If Y and Z were independent, the joint probability $P(Y, Z)$ would have been equal to $P(Y) * P(Z)$, which would've been non-zero, since we find that both $P(Z)$ and $P(Y)$ are non-zero. However, since the joint probability $P(Y, Z) = 0$, Y and Z are not independent.

3 Optimization

Given, optimization problem:

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

(a) Legrange function:
For the given optimization problem, the Lagrange function is

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

(b) KKT conditions:
The KKT conditions for the given optimization problem are:

- $$\frac{\partial L}{\partial x} = 4x + 3y - \lambda_1 x = 0$$
- $$\frac{\partial L}{\partial y} = 3x - \lambda_1 + \lambda_2 = 0$$
- $$\lambda_1(\frac{x^2}{2} + y - 4) = 0$$
- $$\lambda_2(-y + 2) = 0$$

(c) Solving for 4 possibilities formed by each constraint being active or inactive:

- Both constraints active:
  This gives us:

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

  Using the value of y, we get:

$$\begin{aligned} x &= \sqrt{2(4-y)} \\ &= \sqrt{4} \\ &= \pm 2 \end{aligned}$$

  - Considering $x = 2$, from the KKT conditions we get:

$$4x + 3y - \lambda_1 x = 0$$
$$8 + 6 - 2\lambda_1 = 0$$
$$\lambda_1 = 7$$

    Using the values of x, y and λ_1:

$$3x - \lambda_1 + \lambda_2 = 0$$
$$6 - 7 + \lambda_2 = 0$$
$$\lambda_2 = 1$$

    Since the values of $x, y, \lambda_1 and \lambda_2$ satisfy our constraints, we get a candidate point:

$$(x, y) = (2, 2)$$

  - Considering $x = -2$, from the KKT conditions we get:

$$4x + 3y - \lambda_1 x = 0$$
$$-8 + 6 + 2\lambda_1 = 0$$
$$\lambda_1 = 1$$

    Using the values of x, y and λ_1:

$$3x - \lambda_1 + \lambda_2 = 0$$
$$-6 - 1 + \lambda_2 = 0$$
$$\lambda_2 = 7$$

    Since the values of $x, y, \lambda_1 and \lambda_2$ satisfy our constraints, we get a candidate point:

$$(x, y) = (-2, 2)$$

- Constraint 1 active and Constraint 2 inactive:
  This gives us:
$$\frac{x^2}{2} + y = 4, \lambda_1 > 0$$
$$y > 2, \lambda_2 = 0$$
  Using the value of λ_2, from the KKT conditions, we get:
$$3x - \lambda_1 + \lambda_2 = 0$$
$$3x - \lambda_1 = 0$$
$$x = \frac{\lambda_1}{3}$$
  This gives us:
$$y = 4 - \frac{x^2}{2}$$
$$y = 4 - \frac{\lambda_1^2}{18}$$
  Using this to find λ_1, we get:
$$4x + 3y - \lambda_1 x = 0$$
$$\frac{4\lambda_1}{3} + 12 - \frac{\lambda_1^2}{6} - \frac{\lambda_1^2}{3} = 0$$
  Solving for λ_1, we get:
$$\lambda_1 = \frac{4 + 2\sqrt{58}}{3}$$
$$x = 2.137$$
$$y = 1.717$$
  Since the value of y is supposed to be > 2, we have a contradiction.
- Constraint 1 inactive and Constraint 2 active:
  This gives us:
$$\frac{x^2}{2} + y < 4, \lambda_1 = 0$$
$$y = 2, \lambda_2 > 0$$
  Using the value of λ_1 and y, from the KKT conditions, we get:
$$4x + 3y - \lambda_1 x = 0$$
$$x = \frac{-3}{2}$$
  Using the values of x, y and λ_1, we get:
$$3x - \lambda_1 + \lambda_2 = 0$$

$$\lambda_2 = \frac{9}{2}$$

Since the values of $x, y, \lambda_1 and \lambda_2$ satisfy our constraints, we get a candidate point:

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

- Both constraints are inactive:
  This gives us:

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

$$y > 2, \lambda_2 = 0$$

  Using the value of λ_1 and λ_2, from the KKT conditions, we get:

$$3x - \lambda_1 + \lambda_2 = 0$$

$$x = 0$$

  Solving for y. we get:

$$4x + 3y - \lambda_1 x = 0$$

$$y = 0$$

  This does not satisfy our constraint $y > 2$, hence this cannot be considered a candidate point.

(d) All candidate points:
The candidate points for our optimization problem are:

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

(e) Checking for maximality and sufficiency:
To check for maximality, we will evaluate the function $f(x, y)$ at each of our candidate points.

- $$(x, y) = (2, 2)$$
$$\begin{aligned} f(x, y) &= 2x^2 + 3xy \\ &= 2 \times 4 + 3 \times 2 \times 2 \\ &= 8 + 12 \\ &= 20 \end{aligned}$$

- $$(x, y) = (-2, 2)$$
$$\begin{aligned} f(x, y) &= 2x^2 + 3xy \\ &= 2 \times 4 - 3 \times 2 \times 2 \\ &= 8 - 12 \\ &= -4 \end{aligned}$$

- $$(x, y) = (\frac{-3}{2}, 2)$$
$$\begin{aligned} f(x, y) &= 2x^2 + 3xy \\ &= 2 \times \frac{9}{4} - 3 \times \frac{3}{2} \times 2 \\ &= -\frac{9}{2} \end{aligned}$$

We see that the candidate point $(x, y) = (2, 2)$ maximizes our function.
To check for sufficiency, we will see if the Lagrange function $L(x, y)$ is concave at this point. For this candidate point, we had found the values of λ_1 and λ_2 to be:

$$\lambda_1 = 7$$
$$\lambda_2 = 1$$

Using these values, we get:

$$\begin{aligned} L(x, y) &= 2x^2 + 3xy - \lambda_1(\frac{x^2}{2} + y - 4) - \lambda_2(-y + 2) \\ &= 2x^2 + 3xy - 7(\frac{x^2}{2} + y - 4) - (-y + 2) \\ &= 2x^2 + 3xy - 7\frac{x^2}{2} - 7y + 28 + y - 2 \\ &= -\frac{3x^2}{2} + 3xy - 6y + 26 \end{aligned}$$

Since the coefficient of x^2 is negative, we can see that this will be a concave graph. Hence, our candidate point of $(x, y) = (2, 2)$ is sufficiently the optimal solution.

4 Maximum Likelihood

4.1 Discrete Example

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.

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

(a) Likelihood of the result given θ:
Since the probability of picking up either coin for a flip is equal, we have

$$P(A) = P(B) = 0.5$$

Since all the events are independent of each other, we have

$$\begin{aligned} L(R|\theta) &= \prod_i f_{x_i}(x : \theta) \\ &= (\frac{1}{2})^5 \times (1-\theta)^5 \times (\frac{1}{2})^2 \times \theta^2 \times (\frac{1}{2})^3 \times (2\theta)^3 \\ &= (\frac{1}{2})^7 \times (1-\theta)^5 \times \theta^5 \end{aligned}$$

(b) Maximum likelihood estimation for θ:
To find the maximum likelihood estimation of θ, we need to find the value of θ such that:

$$\frac{dL}{d\theta} = 0$$

and

$$\frac{d^2L}{d^2\theta} < 0$$

Calculating $\frac{dL}{d\theta}$ using the product rule of differentiation, we get:

$$\begin{aligned}\frac{dL}{d\theta} &= \frac{d((\frac{1}{2})^7 \times (1-\theta)^5 \times \theta^5)}{d\theta} \\ &= (\frac{1}{2})^7 \times \frac{d((1-\theta)^5 \times \theta^5}{d\theta} \\ &= (\frac{1}{2})^7 \times (-5(1-\theta)^4\theta^5 + 5(1-\theta)^5\theta^4) \\ &= 5 \times (\frac{1}{2})^7 \times \theta^4(1-\theta)^4(1-2\theta)\end{aligned}$$

For $\frac{dL}{d\theta} = 0$, we get the possible values of θ as $\theta = 0$, $\theta = 1$ and $\theta = \frac{1}{2}$.
Now, to find the θ that maximizes the likelihood estimation, we will find the values of $\frac{d^2L}{d^2\theta}$ for each of these values. we find that:

For $\theta = 0$, $\frac{d^2L}{d^2\theta} = 0$
For $\theta = 1$, $\frac{d^2L}{d^2\theta} = 0$
For $\theta = \frac{1}{2}$, $\frac{d^2L}{d^2\theta} = -\frac{5}{2^14}$
Since $\frac{d^2L}{d^2\theta} < 0$ for $\theta = \frac{1}{2}$, the maximum likelihood estimation for θ is 0.5.

4.2 Normal distribution

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) Log likelihood function:

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

The X_n are independent (as the last equality is suggesting) but not identically distributed (they have different means)

Given,

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

We can write,

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

where, $Z_n \sim \text{Normal}\left(0, \sigma^2\right)$. Thus, we get:

$$f_{X_i}(x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta t_i^3)^2}{2\sigma^2}}$$

Using this to calculate log likelihood, we get:

$$\begin{aligned}
\ell\left(\theta; x_1, \ldots, x_N\right) &= \log f_{X_1, \ldots, X_N}\left(x_1, \ldots, x_N; \theta\right) \\
&= \log\left(f_{X_1}\left(x_1; \theta\right) f_{X_2}\left(x_2; \theta\right) \cdots f_{X_N}\left(x_N; \theta\right)\right) \\
&= \log \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta t_i^3)^2}{2\sigma^2}}\right) \\
&= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta t_i^3)^2}{2\sigma^2}}\right) \\
&= N \log \frac{1}{\sqrt{2\pi\sigma^2}} - \sum_{i=1}^{N} \frac{1}{2\sigma^2}((x_i - \theta t_i^3)^2) \\
&= N \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{i=1}^{N}((x_i - \theta t_i^3)^2)
\end{aligned}$$

(b) MLE for θ:
To calculate the maximum likelihood estimation for θ, we need to find the value of θ for which

$$\frac{d\ell\left(\theta\right)}{d\theta} = 0$$

$$\begin{aligned}\frac{d\ell\left(\theta\right)}{d\theta} &= \frac{d(N \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}((x_i - \theta t_i^3)^2))}{d\theta} \\ &= 0 + \frac{1}{2\sigma^2}\sum_{i=1}^{N} 2(x_i - \theta t_i^3)t_i^3\end{aligned}$$

Equating this to 0, we get:

$$\frac{1}{2\sigma^2}\sum_{i=1}^{N} 2(x_i - \theta t_i^3)t_i^3 = 0$$

$$\sum_{i=1}^{N}(x_i - \theta t_i^3)t_i^3 = 0$$

$$\sum_{i=1}^{N} x_i t_i^3 - \sum_{i=1}^{N} \theta t_i^6 = 0$$

$$\theta = \frac{\sum_{i=1}^{N} x_i t_i^3}{\sum_{i=1}^{N} t_i^6}$$

Thus, the maximum likelihood estimation for θ is

$$\theta = \frac{\sum_{i=1}^{N} x_i t_i^3}{\sum_{i=1}^{N} t_i^6}$$

4.3

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) P.D.F of above independent random variables:
The PDF of random variable X is:

$$\begin{aligned} p(X) &= \frac{dP(X)}{dx} \\ &= \begin{cases} \alpha(\frac{x^{\alpha-1}}{\beta^\alpha}), & 0 \leq x \leq \beta \\ 0, & x > \beta \end{cases} \end{aligned}$$

(b) MLEs of α and β:
To calculate the MLEs of α and β, we first find the likelihood function.

$$\begin{aligned} L(X|\alpha,\beta) &= \prod_{i=1}^{N} f_{x_i}(x:\alpha,\beta) \\ &= \prod_{i=1}^{N} \alpha(\frac{x_i^{\alpha-1}}{\beta^{\alpha}}) \\ &= \frac{\alpha^N}{\beta^{N\alpha}} \prod_{i=1}^{N} x_i^{\alpha-1} \end{aligned}$$

From this, we get the log likelihood function to be:

$$\begin{aligned} l(X) &= \log L(X) \\ &= \log(\frac{\alpha^N}{\beta^{N\alpha}} \prod_{i=1}^{N} x_i^{\alpha-1}) \\ &= \log \frac{\alpha^N}{\beta^{N\alpha}} + \log \prod_{i=1}^{N} x_i^{\alpha-1} \\ &= \log \frac{\alpha^N}{\beta^{N\alpha}} + \sum_{i=1}^{N} \log x_i^{\alpha-1} \\ &= N\log\alpha - N\alpha\log\beta + (\alpha-1)\sum_{i=1}^{N}\log x_i \end{aligned}$$

Now, to find the MLE of α,

$$\frac{dl(X)}{d\alpha} = 0$$

Substituting the value of l(X), we get:

$$\begin{aligned} \frac{dl(X)}{d\alpha} &= \frac{d(N\log\alpha - N\alpha\log\beta + (\alpha-1)\sum_{i=1}^{N}\log x_i)}{d\alpha} \\ &= \frac{N}{\alpha} - N\log\beta + \sum_{i=1}^{N}\log x_i \end{aligned}$$

Equating this to 0, we get,

$$\frac{N}{\alpha} - N\log\beta + \sum_{i=1}^{N}\log x_i = 0$$

$$\frac{N}{\alpha} = N\log\beta - \sum_{i=1}^{N}\log x_i$$

$$\alpha = \frac{N}{N\log\beta - \sum_{i=1}^{N}\log x_i}$$

For the MLE for β, if we look at the likelihood function,

$$L(X|\alpha,\beta) = \frac{\alpha^N}{\beta^{N\alpha}}\prod_{i=1}^{N} x_i^{\alpha-1}$$

We notice that β is in the denominator, hence, the likelihood is maximum when β is minimum. Since x ranges from $0 \le x \le \beta$, β is minimum when x is maximum. Thus, the MLE for β is:

$$\beta = X_n$$

Therefore, the MLEs for α and β:

$$\alpha = \frac{N}{N\log\beta - \sum_{i=1}^{N}\log x_i}$$

$$\beta = X_n$$

5 Information Theory

5.1 Marginal Distribution

Given joint probability distribution of two binary random variables X and Y:

(a) Marginal distribution of X and Y, respectively.
From the given joint distribution of random variables X and Y, the marginal distribution of X can be calculated using the formula:

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

Thus, the marginal distribution of X is:

x	$f_X(x)$
0	2/3
1	1/3

Similarly, the marginal distribution of Y can be calculated using the formula:

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Thus, the marginal distribution of Y is:

y	$f_Y(y)$
1	1/3
2	2/3

(b) Mutual information for the joint probability distribution:
The mutual information between 2 random variables can be calculated using the formula

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \\ &= \sum_{x,y} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)} \end{aligned}$$

Substituting the values from the marginal distribution tables and the joint probability distribution table, we get:

$$\begin{aligned} I(X,Y) &= \frac{1}{3}\log_2\frac{3}{2} + \frac{1}{3}\log_2\frac{3}{4} + \frac{1}{3}\log_2\frac{3}{2} \\ &= \frac{1}{3}(\log_2\frac{3}{2} + \log_2\frac{3}{4} + \log_2\frac{3}{2}) \\ &= \frac{1}{3}\log_2\frac{27}{16} \end{aligned}$$

5.2 Mutual Information and Entropy

Given the dataset,

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) Entropy $H(Y)$:
To find the entropy of Y, we will first find out the marginal distribution of Y using the formula

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Thus, the marginal distribution of Y is:

y	$f_Y(y)$
yes	4/7
no	3/7

Now, the entropy of Y can be calculated as:

$$\begin{aligned} H(Y) &= \sum p(Y)I(Y) \\ &= -\sum_k p(Y = k) \log_2 p(Y = k) \\ &= -(\frac{4}{7} \log_2 \frac{4}{7} + \frac{3}{7} \log_2 \frac{3}{7}) \\ &= -(-0.463 - 0.523) \\ &= 0.986 \end{aligned}$$

(b) Conditional entropy $H(Y|x_1)$, $H(Y|x_4)$, respectively:

- Calculating $H(Y|x_1)$:

$$\begin{aligned} H(Y|x_1) &= \sum_x p(x_1)H(Y|X=x_1) \\ &= \sum_{x,y} p(x_1,y_i)\log_2 \frac{p(x_1)}{p(x_1)p(y)} \\ &= \frac{1}{14}\log_2 5 + \frac{4}{14}\log_2 \frac{5}{4} + \frac{3}{14}\log_2 \frac{5}{3} + \frac{2}{14}\log_2 \frac{5}{2} \\ &= \frac{1}{14}(\log_2 5 + 4\log_2 \frac{5}{4} + 3\log_2 \frac{5}{3} + 2\log_2 \frac{5}{2}) \\ &= 0.604 \end{aligned}$$

- Calculating $H(Y|x_4)$:

$$\begin{aligned} H(Y|x_4) &= \sum_x p(x_4)H(Y|X=x_4) \\ &= \sum_{x,y} p(x_4,y_i)\log_2 \frac{p(x_4)}{p(x_4)p(y)} \\ &= \frac{5}{14}\log_2 \frac{8}{5} + \frac{3}{14}\log_2 2 + \frac{3}{14}\log_2 \frac{8}{3} + \frac{3}{14}\log_2 2 \\ &= \frac{1}{14}(6 + 5\log_2 \frac{8}{5} + 3\log_2 \frac{8}{3}) \\ &= 0.973 \end{aligned}$$

(c) Mutual information $I(x_1, Y)$ and $I(x_4, Y)$:

- Calculating mutual information $I(x_1, Y)$:

  $$I(x_1, Y) = H(Y) - H(Y|x_1)$$

  Substituting the values of $H(Y)$ and $H(Y_{x1})$ from parts a) and b) respectively, we get:

  $$\begin{aligned} I(x_1, Y) &= H(Y) - H(Y|x_1) \\ &= -(\frac{4}{7}\log_2\frac{4}{7} + \frac{3}{7}\log_2\frac{3}{7}) - \frac{1}{14}(\log_2 5 + 4\log_2\frac{5}{4} + 3\log_2\frac{5}{3} + 2\log_2\frac{5}{2}) \\ &= -\frac{1}{14}(8\log_2\frac{4}{7} + 6\log_2\frac{3}{7} + \log_2 5 + 4\log_2\frac{5}{4} + 3\log_2\frac{5}{3} + 2\log_2\frac{5}{2}) \\ &= 0.382 \end{aligned}$$

- Calculating mutual information $I(x_4, Y)$:

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

  Substituting the values of $H(Y)$ and $H(Y_{x4})$ from parts a) and b) respectively, we get:

  $$\begin{aligned} I(x_1, Y) &= H(Y) - H(Y|x_1) \\ &= -(\frac{4}{7}\log_2\frac{4}{7} + \frac{3}{7}\log_2\frac{3}{7}) - \frac{1}{14}(6 + 5\log_2\frac{8}{5} + 3\log_2\frac{8}{3}) \\ &= -\frac{1}{14}(8\log_2\frac{4}{7} + 6\log_2\frac{3}{7} + 6 + 5\log_2\frac{8}{5} + 3\log_2\frac{8}{3}) \\ &= 0.013 \end{aligned}$$

  Since, $x1$ has higher mutual information, it is more informative.

(d) Joint entropy $H(Y, x_3)$:
Joint entropy $H(Y, x_3)$ can be calculated by using the formula:

$$\begin{aligned}
H(Y, x_3) &= \sum_{y, x_3} P(Y, X_3) \log_2 \frac{1}{P(Y, x_3)} \\
&= -\sum_{y, x_3} P(Y, X_3) \log_2 P(Y, x_3) \\
&= -(\frac{4}{14} \log_2 \frac{4}{14} + \frac{3}{14} \log_2 \frac{3}{14} + \frac{2}{14} \log_2 \frac{2}{14} + \frac{5}{14} \log_2 \frac{5}{14}) \\
&= -\frac{1}{14}(4 \log_2 \frac{4}{14} + 3 \log_2 \frac{3}{14} + 2 \log_2 \frac{2}{14} + 5 \log_2 \frac{5}{14}) \\
&= 1.924
\end{aligned}$$

5.3 Entropy Proofs

(a) Suppose X and Y are independent. Proof that $H(X|Y) = H(X)$:
Given, X and Y are independent. We know that:

$$H(X|Y) = -\sum_x \sum_y p(x,y) \log p(x|y)$$

Since X and Y are independent,

$$p(x,y) = p(x)p(y)$$

and

$$p(x|y) = p(x)$$

Substituting these values in the equation above:

$$\begin{aligned} H(X|Y) &= -\sum_x \sum_y p(x,y) \log p(x|y) \\ &= -\sum_x \sum_y p(x)p(y) \log p(x) \\ &= -\sum_x (\sum_y p(y)) p(x) \log p(x) \end{aligned}$$

Sum over all probabilities of y = 1. Therefore, the above equation becomes

$$\begin{aligned} H(X|Y) &= -\sum_x (\sum_y p(y)) p(x) \log p(x) \\ &= -\sum_x p(x) \log p(x) \\ &= H(X) \end{aligned}$$

Hence, proved!

(b) Suppose X and Y are independent. Proof that $H(X,Y) = H(X)+H(Y)$: Given, X and Y are independent. We know that:

$$H(X,Y) = -\sum_x \sum_y p(x,y) \log p(x,y)$$

Since X and Y are independent,

$$p(x,y) = p(x)p(y)$$

Substituting this value in the equation above:

$$\begin{aligned} H(X,Y) &= -\sum_x \sum_y p(x,y) \log p(x,y) \\ &= -\sum_x \sum_y p(x)p(y) \log p(x)p(y) \\ &= -\sum_x \sum_y p(x)p(y)(\log p(x) + \log p(y)) \\ &= -\sum_x \sum_y p(x)p(y) \log p(x) - \sum_x \sum_y p(x)p(y) \log p(y) \\ &= -\sum_x (\sum_y p(y))p(x) logp(x) - \sum_y (\sum_x p(x))p(y) \log p(y) \end{aligned}$$

Sum over all probabilities of y = 1. Similarly, sum of all probabilities of x = 1. Therefore, the above equation becomes

$$\begin{aligned} H(X,Y) &= -\sum_x (\sum_y p(y))p(x) logp(x) - \sum_y (\sum_x p(x))p(y) \log p(y) \\ &= -\sum_x p(x) \log p(x) - \sum_y p(y) \log p(y) \\ &= H(X) + H(Y) \end{aligned}$$

Hence, proved!

(c) Proof that the mutual information is symmetric, i.e., $I(X, Y) = I(Y, X)$ and $x_i \in X, y_i \in Y$:
We know that mutual information can be calculated as:

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

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

We also know that,

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

From the above equation, we get,

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

Using this value of $H(X|Y)$ in the calculation for mutual information, we get,

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) \\ &= H(X) - H(X) - H(Y|X) + H(Y) \\ &= H(Y) - H(Y|X) \\ &= I(Y, X) \end{aligned}$$

Hence, proved!

6

(a) Given, a random variable X has a Poisson distribution with mean 8, the expectation E$[(X+2)^2]$:
We know that the variance and mean of a Poisson distribution are equal. Hence, we have:

$$Var(X) = E[X] = 8$$

Now,

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

Using the value of $E[X]$, we get:

$$\begin{aligned} E[(X+2)^2] &= E[X^2] + E[4] + 4E[X] \\ &= E[X^2] + 4 + 4 \times 8 \\ &= 36 + E[X^2] \end{aligned}$$

Now, we know that:

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

Substituting the values of $Var(X)$ and $E[X]$, we get,

$$8 = E[X^2] - 8^2$$

$$E[X^2] = 72$$

Thus, we have,

$$\begin{aligned} E[(X+2)^2] &= 36 + E[X^2] \\ &= 36 + 72 \\ &= 108 \end{aligned}$$

(b) A person decides to toss a fair coin repeatedly until he gets a head. He will make at most 3 tosses. Random variable Y denotes the number of heads. Variance of Y:
Given, the person tosses a fair coin repeatedly until he gets a head. Therefore, the possible outcomes of this event are (H), (TH), (TTH) and (TTT). Since the coin is fair, we can calculate the probability of each of these events as:

$$P((H)) = 0.5$$

$$P((TH)) = 0.5 \times 0.5$$

$$P((TTH)) = 0.5 \times 0.5 \times 0.5$$

$$P((TTT)) = 0.5 \times 0.5 \times 0.5$$

Now, Y denotes the number of heads. From the possible outcomes, we can see that Y can either be 0 or 1. Thus,

$$\begin{aligned} E[Y] &= 0 \times p(Y = 0) + 1 \times p(Y = 1) \\ &= 0 + (0.5 + 0.5 \times 0.5 + 0.5 \times 0.5 \times 0.5) \\ &= 0.875 \end{aligned}$$

Similarly,

$$\begin{aligned} E[Y^2] &= 0^2 \times p(Y = 0) + 1^2 \times p(Y = 1) \\ &= 0 + (0.5 + 0.5 \times 0.5 + 0.5 \times 0.5 \times 0.5) \\ &= 0.875 \end{aligned}$$

Now, variance of Y can be calculated as,

$$\begin{aligned} var(Y) &= E[Y^2] - E[Y]^2 \\ &= 0.875 - 0.875^2 \\ &= 0.109375 \end{aligned}$$

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

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

Probability P(X+Y $\le$ 1):
We can calculate P(X+Y $\le$ 1) as:

$$\begin{aligned} P(X+Y \le 1) &= \int_0^1 \int_0^{1-x} (x+y) dy dx \\ &= \int_0^1 [yx + \frac{y^2}{2}]_0^{1-x} dx \\ &= \int_0^1 ((1-x)x + \frac{(1-x)^2}{2} - 0) dx \\ &= \int_0^1 (x - x^2 + \frac{(1+x^2-2x)}{2}) dx \\ &= \frac{1}{2} \int_0^1 (2x - 2x^2 + 1 + x^2 - 2x) dx \\ &= \frac{1}{2} \int_0^1 (-x^2 + 1) dx \\ &= \frac{1}{2} \times [-\frac{x^3}{3} + x]_0^1 \\ &= \frac{1}{2} \times (\frac{1}{3} + 1 - 0) \\ &= \frac{1}{2} \times (\frac{2}{3}) \\ &= \frac{1}{3} \end{aligned}$$