Machine Learning (ML) in Hindi

Week-2 Assignment

Total points: 21

Instructions

- Each question is a 1-point multiple choice question (MCQ) with only one correct answer unless specified otherwise.
- Multiple select questions (MSQ) are 2-point questions with one or more than one answer correct.
- Numerical-answer type questions carry 2-points and have an exact numeric answer. Try to be as precise as possible.
- For programming questions, use the default parameter settings unless specified otherwise.

1. [MSQ] Given the following data, select the correct statement(s).

X	Y
2	1
4	3
6	8
8	11
10	12

   a. The magnitude of the difference between the means of X & Y is 1
   b. The magnitude of the difference between the variances of X & Y is 10
   c. The magnitude of the covariance of X & Y is 12
   d. The covariance matrix of X & Y is a Positive Semi-Definite (PSD) Matrix
   e. The correlation coefficient for this dataset is 0.5

Solution: (a), (c), (d)

- Mean,

$$\mu_X = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{5} X_i}{5} = 6$$

  &

$$\mu_Y = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{5} Y_i}{5} = 7$$

  Then, the magnitude of the difference between the means is given by

$$|\mu_X - \mu_Y| = |6 - 7| = 1$$

- Variance,

$$\sigma_X^2 = \frac{\sum_{i=1}^{n}(X_i - \mu_X)^2}{n} = \frac{\sum_{i=1}^{5}(X_i - \mu_X)^2}{5} = \frac{16+4+0+4+16}{5} = 8$$

&

$$\sigma_Y^2 = \frac{\sum_{i=1}^{n}(Y_i - \mu_Y)^2}{n} = \frac{\sum_{i=1}^{5}(Y_i - \mu_Y)^2}{5} = \frac{36+16+1+16+25}{5} = 18.8$$

Then, the magnitude of the difference between the variances is given by

$$|\sigma_X^2 - \sigma_Y^2| = |8 - 18.8| = 10.8$$

- Covariance,

$$\mathrm{Cov}(X,Y) = \sum_{i=1}^{n}\frac{(X_i - \mu_X)(Y_i - \mu_Y)}{n} = \sum_{i=1}^{5}\frac{(X_{i-\mu_X})(Y_i - \mu_Y)}{5}$$

$$= \frac{(2-6)(1-7) + (4-6)(3-7) + (6-6)(8-7) + (8-6)(11-7) + (10-6)(12-7)}{5}$$

$$= \frac{(-4)(-6) + (-2)(-4) + 0 + (2)(4) + (4)(5)}{5} = \frac{24+8+8+20}{5} = \frac{60}{5} = 12$$

- We know that, a covariance matrix is at least a PSD matrix.
  Here, it is given by

$$\mathbf{K}(X,Y) = \begin{bmatrix} \sigma_X^2 & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 12 & 18.8 \end{bmatrix}$$

  This is a symmetric matrix [$\mathbf{K}^T = \mathbf{K}$]. A symmetric matrix is PSD if and only if all its eigenvalues are non-negative.

$$\det(\mathbf{K} - \lambda\mathbf{I}) = \mathbf{0}$$
$$(8-\lambda)(18.8-\lambda) - 144 = 0$$
$$\lambda^2 - 26.8\lambda + 6.4 = 0$$
$$\lambda = 26.56, 0.24$$

  Since both the eigenvalues are non-negative, $\mathbf{K}(X,Y)$ is a PSD matrix. Hence Proved.

- Correlation Coefficient,

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}} = \frac{12}{\sqrt{8}\sqrt{18.8}} = 0.98$$

2. [Numerical Type] Let X and Y be two normally distributed independent random variable with unit variance and zero mean. Also, let $M = 1 + X + XY^2$ and $N = 1 + X$. Calculate the value of covariance between M & N.

Solution: 2

$$\mathrm{Conv}(M,N) = \mathrm{Cov}(1 + X + XY^2, 1 + X)$$
$$= E[(1 + X + XY^2)(1 + X)] - E[1 + X + XY^2]E[1 + X]$$
$$= E[1 + X + XY^2 + X + X^2 + X^2Y^2] - (E[1] + E[X] + E[XY^2])(E[1] + E[X])$$

We know that,

- $E[1] = 1$
- Zero Mean: $E[X] = 0, E[Y] = 0$
- Independent Variables: $E[XY] = E[X]E[Y] = 0$

Then,

$$\text{Conv(M, N)} = 1 + 0 + \text{E[X]E[Y}^2\text{]} + 0 + \text{E}[X^2] + E[X^2]E[Y^2] - (1 + 0 + 0)(1 + 0)$$
$$= 1 + 0 + E[X^2] + E[X^2]E[Y^2] - 1$$

We know that, variance $E[X^2] = E[Y^2] = 1$

Then,

$$\text{Conv(M, N)} = 1 + 1 = 2$$

3. Consider modelling the response variable in terms of the features x_i and the parameters θ_i. Which of the following can NOT be considered a linear regression model?
   a. $y = \theta_0 + \sum_{i=1}^{n} \theta_i \cos x_i$
   b. $y = \theta_0 + \sum_{i=1}^{n} \theta_i x_i^2$
   c. $y = \theta_0 + \sum_{i=1}^{n} \theta_i \theta_{i-1} x_i$
   d. $y = \theta_0 + \sum_{i=1}^{n} \theta_i x_i$

Solution: (c)

Since output y is dependent on $\theta_i \theta_{i-1}$, It is not a linear model.

4. [MSQ] Consider the regression model $y = \boldsymbol{\theta}^T \boldsymbol{x} + \varepsilon$, where y is the response variable, $\boldsymbol{x}$ is a 10-element vector of features, $\boldsymbol{\theta}$ is a 10-element vector of parameters corresponding to each feature, and ε is the model error. Note that we have considered the bias parameter θ_0=0. The model prediction is defined as $\hat{y} = \boldsymbol{\theta}^T \boldsymbol{x}$. Here, subscript to a vector indicates its elements (e.g., x_2 represents 2nd element of the vector x). Identify the FALSE statement(s). [⇒ refers to 'implies']
   a. If $\theta_1 = 0 \Longrightarrow \hat{y}$ does not depend on the first feature.
   b. If $\theta_5 > 0$ and $x_5 > 0$, $\Rightarrow \hat{y}$ must be greater than zero.
   c. If $\theta_9 = 0 \Longrightarrow$ increase in x_9 will lead to a decrease in $\hat{y}$ [Keeping everything else same].
   d. ε can be absorbed in $\boldsymbol{\theta}$ after making relevant changes in $\boldsymbol{x}$ & $\boldsymbol{\theta}$

Solution: (b), (c), (d)

- θ_5 & x_5 alone cannot dictate the value of $\hat{y}$. E.g., ε might become too large and negative.
- If $\theta_9 = 0$, increasing x_9 will not impact $\hat{y}$ at all.
- ε is a i.i.d. random variable that will change for different data samples, but $\boldsymbol{\theta}$ is a constant for all data samples.

5. A frozen dessert startup sells k different types of frozen desserts and aims to become a unicorn soon. In order to meet its overall profit targets, the startup is considering a change in the price of one of its articles. The startup hires an ML expert who develops a regression model that predicts the total profit when the product prices are changed, given by $\hat{d} = \boldsymbol{\theta}^T \boldsymbol{x} + d$, where $\hat{d}$ is the predicted profit after the increase in price, $\boldsymbol{\theta}$ is the learned model parameters (based on the past data of the impact on the company's profit with altered article prices over the

years), d is the current profit, and $\boldsymbol{x}$ denotes the k-elements vector representing a fractional change in the prices of the articles, $x_i = \frac{p_i^{new} - p_i}{p_i}$, where, p_i is the current (positive) price of the i^{th} product and p_i^{new} is the new price of the i^{th} product. What does $\theta_2 < 0$ imply?

a. Profit increases with an increase in the price of the 2nd product.
b. Profit decreases with an increase in the price of the 2nd product.
c. The profit of all the products is negatively impacted.
d. This information is insufficient for any relevant conclusions.

Solution: (b)

6. For an unbiased linear regression or the ordinary least squares (OLS) regression, the error term should be uncorrelated with the _____ and should maintain a zero _____ . Select the correct combination that fills the two blanks.

a. Independent variable, mean
b. Dependent variable, mean
c. Independent variable, standard deviation
d. Dependent variable, standard deviation

Solution: (a)

- Suppose an independent variable is correlated with the error term. In that case, one can use the independent variable to predict the error term, which violates the notion that the error term represents unpredictable random error.
- The error term accounts for the variation in the dependent variable that the independent variables do not explain. Random chance should determine the values of the error term. For the model to be unbiased, the average value of the error term must be equal to zero.

7. [MSQ] Which of the following statement(s) about the closed-form solution for Linear Regression is/are TRUE?

a. It is computationally expensive for large datasets.
b. It is only applicable to linearly separable data.
c. It is a global optimization problem.
d. It is robust to changes in the dataset.

Solution: (a), (c)

8. [Numerical Type] In a game show, a G units long conveyer belt moves a water-filled balloon from one end to another. Players are given access to two sets of needles placed on the top of the conveyer belt at locations $\left[\frac{G}{6}, \frac{G}{3}\right]$ & $\left[\frac{2G}{3}, G\right]$ that puncture the balloon placed at that location by the push of a button. Player receives 100 points if the balloon bursts. Let g be the location of the balloon at the time the button is pressed. The belt is moving at a constant speed. Let us define a discrete (binary) random variable $M \in \{0, 100\}$ describing the event of pressing the button by the player & receiving points, as shown:

$$M(g) = \begin{cases} 100, & \text{if } g \in \left[\frac{G}{6}, \frac{G}{3}\right] \cap \left[\frac{2G}{3}, G\right] \\ 0, & \text{otherwise} \end{cases}$$

What is the probability that the player did not receive any point in the current push of the button?

Solution: 0.5

$$P[M(g) = 100] = \frac{\left\{\frac{G}{3} - \frac{G}{6} + G - \frac{2G}{3}\right\}}{\{G\}} = \frac{\{2 - 1 + 6 - 4\}}{6} = \frac{1}{2}$$

$$P[M(g) = 0] = 1 - 0.5 = 0.5$$

9. [Numerical Type] Given the following data, what is the sum of the regression parameters (θ_0 & θ_1)? Solve using closed form solution.

X	Y
0	7
1	9
-2	3
3	13
-3	1

Solution: 7+2=9

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -2 \\ 1 & 3 \\ 1 & -3 \end{bmatrix} \;\&\; y = \begin{bmatrix} 7 \\ 9 \\ 3 \\ 13 \\ 1 \end{bmatrix}$$

We know that,

$$\theta = (X^T X)^{-1} X^T y = \frac{1}{114} \begin{bmatrix} 23 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 33 \\ 39 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}$$

10. [MSQ] Which of the following properties are assumed for errors in a linear regression?
    a. Multicollinearity
    b. Independence
    c. Homoscedasticity
    d. Normality
    e. Identical distribution
    f. Exponential distribution

Solution: (b), (c), (d), (e)

11. What is the output of the following Python code?

```
from sklearn.linear_model import LinearRegression
X = [[1], [2], [3], [4], [5]]
y = [2, 4, 8, 16, 24]
reg = LinearRegression()
reg.fit(X, y)
print(reg.predict([[6]]))
```

a. 30.5
b. 34.1
c. 23.78
d. 27.6

Solution: (d)

The code defines a simple linear regression model using the LinearRegression class from the sklearn.linear_model module, with input X and output y. The model is then trained using the fit() method. Finally, the code prints the predicted output for a new input value of 6, which will be 27.6.

12. Given the following code snippet, what are the (approximate) regression coefficients for the learnt model?

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[1, 3], [2, 4], [5, 7]])
y = np.array([2, 3, 6])
reg = LinearRegression()
reg.fit(X, y)
```

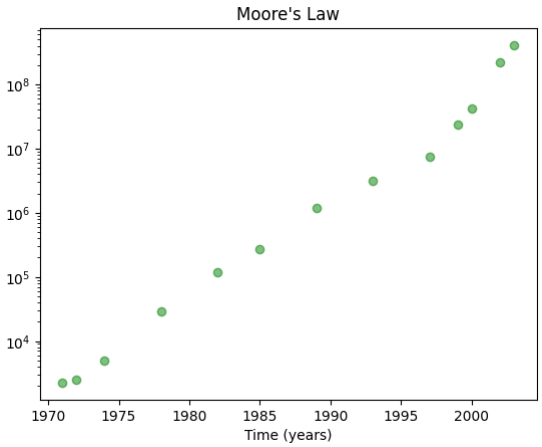
a. [-1.5, 1.5]
b. [0.5 ,0.5]
c. [-0.5, 0.5]
d. [1,2]

Solution: (b)

The output of print(reg.coef_) is [0.50972763 0.49027237]~ [0.5, 0.5]

13. Verify Moore's Law using Python Code. According to Moore's Law, given by Intel co-founder Gordon Moore, the number of transistors per IC roughly doubles every second year. The data and the corresponding visualization (y-axis depicts the

Year	Transistors
1971	2,250
1972	2,500
1974	5,000
1978	29,000
1982	120,000
1985	275,000
1989	1,180,000
1993	3,100,000
1997	7,500,000
1999	24,000,000
2000	42,000,000
2002	220,000,000
2003	410,000,000



number of transistors) have been provided.

Use the data in the given table about the year-wise list of the number of transistors contained in an IC produced that year, to learn a linear regression model and predict the number of transistors in the IC produced in the year 2010. The value will be:

a. 1.94×10^9
b. 4.84×10^9
c. 7.13×10^9
d. 9.67×10^9

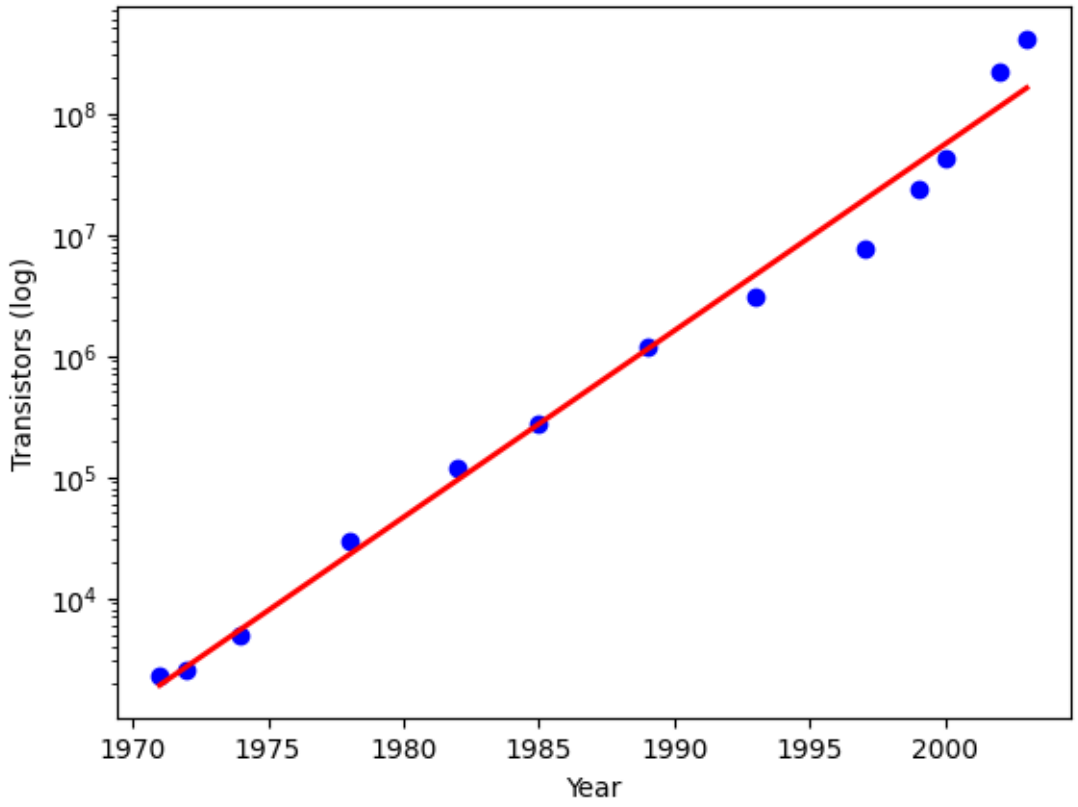
[Note: In case your answer doesn't match any option, select the nearest option]

Solution: (a)

```
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

X = np.array([[1971], [1972],[1974],
[1978],[1982],[1985],[1989],[1993],[1997],[1999],[2000],[2002],[2003]])
y = np.array([2250, 2500, 5000, 29000, 120000, 275000, 1180000, 3100000, 7500000, 24000000,
42000000, 220000000, 410000000])

model = LinearRegression()
model.fit(X,np.log(y))

# Plot the data points
plt.scatter(X, np.log(y), alpha=0.5, color='Green')
plt.yscale("log")
plt.xlabel('Time (years)')
```

```
plt.ylabel('Transistors (log)')
plt.title('Moore\'s Law')
plt.show()
```

```
y_pred = model.predict(X)

plt.yscale("log")
plt.scatter(X, y, color='blue')
plt.plot(X, np.exp(y_pred), color='red', linewidth=2)
plt.xlabel('Time (Years)')
plt.ylabel('Transistors (log)')
plt.show()
```




```
np.exp(model.predict([[2010]]))
```

```
array([1.9333916e+09])
```

14. Given the following code snippet, what does the score function of the LinearRegression() class compute?

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[1, 3], [2, 4], [5, 7]])
y = np.array([2, 3, 6])
reg = LinearRegression()
reg.fit(X, y)
print(reg.score(X, y))
```

a. It computes the mean squared error of the linear regression model.
b. It computes the R-squared score of the linear regression model.
c. It computes the correlation coefficient between the input and output variables.
d. It computes the coefficient of variation of the linear regression model.

Solution: (b)