CAT-1 (Solution Sheet)

1. A small business tracks the number of sales transactions it makes each day. The data for the number of sales per day over a week is provided in the table below:

Day	Number of Sales (X)	Probability (P(X))
Monday	10	0.2
Tuesday	15	0.25
Wednesday	20	0.3
Thursday	25	0.15
Friday	30	0.1

   a) Calculate the expected number of sales per day.
   b) Calculate the variance of the number of sales per day.
   c) Based on the expected sales, what is the total expected number of sales over the week?

Solution:

Part (a): Calculate the Expected Number of Sales per Day

The expected value E(X) is calculated using the formula: $E(X) = \Sigma (X_i \times P(X_i))$

Using the data from the table:

$E(X) = 10 \times 0.2 + 15 \times 0.25 + 20 \times 0.3 + 25 \times 0.15 + 30 \times 0.1 = 18.5$

So, the expected number of sales per day is 18.5.

Part (b): Calculate the Variance of the Number of Sales per Day

The variance Var(X) is calculated using the formula:

$Var(X) = \Sigma [(X_i - E(X))^2 \times P(X_i)]$

First, calculate $(X_i - E(X))^2$ for each X_i:

- For X = 10: (10 - 18.5)^2 = 72.25
- For X = 15: (15 - 18.5)^2 = 12.25
- For X = 20: (20 - 18.5)^2 = 2.25
- For X = 25: (25 - 18.5)^2 = 42.25
- For X = 30: (30 - 18.5)^2 = 132.25

Now, calculate the variance:

Var(X) = 72.25 × 0.2 + 12.25 × 0.25 + 2.25 × 0.3 + 42.25 × 0.15 + 132.25 × 0.1 = 37.75

So, the variance of the number of sales per day is 37.75.

Part (c): Total Expected Number of Sales Over the Week

The total expected number of sales over the week can be calculated by multiplying the expected number of sales per day by the number of days (5 days):

Total Expected Sales = 5 × E(X) = 5 × 18.5 = 92.5

So, the total expected number of sales over the week is 92.5.

2. Given the following matrix A:

$$A = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \\ 3 & 7 & 9 \end{bmatrix}$$

a) Find the rank of the matrix A.
b) Explain the significance of the rank in the context of linear algebra.

Solution:

Part (a): Find the Rank of the Matrix Using Gaussian Elimination

To find the rank of the matrix A, we perform Gaussian elimination to reduce the matrix to its row echelon form (REF). The rank of the matrix is the number of non-zero rows in this form.

Start with the matrix A:

A = [[2, 4, 6], [1, 3, 5], [3, 7, 9]]

1. First step: Eliminate the first column below the first row by subtracting appropriate multiples of the first row from the other rows.

- Subtract (1/2) × (Row 1) from Row 2:

$R_2 \rightarrow R_2 - (1/2)R_1 = [1, 3, 5] - (1/2) [2, 4, 6] = [0, 1, 2]$

- Subtract (3/2) × (Row 1) from Row 3:

$R_3 \rightarrow R_3 - (3/2)R_1 = [3, 7, 9] - (3/2) [2, 4, 6] = [0, 1, 0]$

Now the matrix looks like this:

A = [[2, 4, 6], [0, 1, 2], [0, 1, 0]]

2. Second step: Eliminate the second column below the second row by subtracting Row 2 from Row 3.

- Subtract Row 2 from Row 3:

$R_3 \rightarrow R_3 - R_2 = [0, 1, 0] - [0, 1, 2] = [0, 0, -2]$

The matrix is now in row echelon form:

A = [[2, 4, 6], [0, 1, 2], [0, 0, -2]]

Since all three rows in this matrix are non-zero, the rank of the matrix A is 3.

Part (b): Significance of the Rank

The rank of a matrix is a fundamental concept in linear algebra. It represents the maximum number of linearly independent rows or columns in the matrix. For a matrix A of rank r:

- Dimension of the Column Space: The rank tells us the dimension of the column space (also known as the range) of the matrix. This means that the matrix A can span an r-dimensional space.
- Solutions to Linear Systems: The rank of a matrix is directly related to the solutions of a system of linear equations. If the rank of the coefficient matrix equals the rank of the

augmented matrix, the system is consistent, meaning it has at least one solution.

In this case, the matrix A has a rank of 3, which means its columns (or rows) span a 3-dimensional space. This also indicates that any system of linear equations involving this matrix as a coefficient matrix will have solutions corresponding to this rank.

Thus, the rank of matrix A is 3.

3. A company wants to optimize its production cost. The cost function C(x), where x is the number of units produced, is given by:

$$C(x) = x^2 - 10x + 25$$

   a) Find the number of units x that minimizes the production cost using gradient descent.
   b) Perform gradient descent with an initial guess of $x_0 = 5$ and a learning rate $\alpha = 0.2$. Perform three iterations and provide the updated values of x after each iteration.
   c) Explain how optimizing the number of units produced using gradient descent can lead to cost savings for the company. Discuss the importance of choosing an appropriate learning rate.

Solution:

Part (a): Objective

We aim to find the number of units x that minimizes the production cost $C(x) = x^2 - 10x + 25$.

Part (b): Gradient Descent

Gradient descent updates the value of x iteratively using the following rule:

$x_{ne}w = x_{ol}d - \alpha\ dC(x)/dx$

1. Compute the gradient dC(x)/dx:

$dC(x)/dx = 2x - 10$

2. Initial guess: $x_0 = 5$

3. Iteration 1:

$x_1 = x_0 - \alpha \times dC(x_0)/dx = 5 - 0.2 \times (2(5) - 10) = 5$

(Since the gradient is zero at $x = 5$, the value of x remains unchanged in this iteration.)

4. Iteration 2:

$x_2 = x_1 - \alpha \times dC(x_1)/dx = 5 - 0.2 \times (2(5) - 10) = 5$

5. Iteration 3:

$x_3 = x_2 - \alpha \times dC(x_2)/dx = 5 - 0.2 \times (2(5) - 10) = 5$

Updated Values After Three Iterations:

- $x_1 = 5$
- $x_2 = 5$
- $x_3 = 5$

Observation: As the starting point is chosen exactly at the optimal point, during the subsequent iterations the points do not move.

Part (c): Practical Application and Importance of Learning Rate

- Practical Application: Optimizing the number of units produced using gradient descent helps the company minimize production costs. By determining the optimal number of units x to produce, the company can ensure that it is operating at a cost-effective level, reducing unnecessary expenses and maximizing profits.

- Importance of Learning Rate: In this example, the gradient at the initial guess $x_0 = 5$ is zero, meaning the algorithm has already reached the optimal solution. The learning rate α plays a crucial role in how quickly and effectively the algorithm converges to the minimum cost. If α is too large, the algorithm may overshoot the minimum; if it's too small, the algorithm may take longer to converge.

Final Answer: After three iterations, the optimal number of units to produce remains at x

= 5. The company should produce 5 units to minimize its production cost, ensuring cost efficiency and profitability.

4. The table below provides information related to different houses in terms of square footage (X) and the actual price (Y).

House ID	Square Footage (X)	Actual Price (Y)
1	1000	$150,000
2	1500	$200,000
3	2000	$250,000
4	2500	$300,000
5	3000	$350,000

The task is to predict the price of a house with 2250 square feet using Linear Regression and KNN Regression (with different values of k). In other words, fill up the following table.

Model	k	Predicted Price for 2250 sq. ft. House
Linear Regression	**-**	
KNN Regression	**1**	
KNN Regression	**2**	
KNN Regression	**3**	

Solution:

Using Linear Regression and KNN Regression with different values of k, predict the price for a house with 2250 square feet.

Model	k	Predicted Price for 2250 sq. ft. House
Linear Regression	**-**	$275,000
KNN Regression	**1**	$250,000

KNN Regression	**2**	$275,000
KNN Regression	**3**	$250,000

Explanation:

1. **Linear Regression**: This model provides a global prediction by fitting a straight line through the data points. The predicted price for 2250 sq. ft. is $275,000.
2. **KNN Regression with k=1**: The nearest neighbor to 2250 sq. ft. is House 3 (2000 sq. ft.), which has a price of $250,000. So, the predicted price is $250,000.
3. **KNN Regression with k=2**: The two nearest neighbors are House 3 (2000 sq. ft.) and House 4 (2500 sq. ft.). The predicted price is the average of their prices: (250,000 + 300,000) / 2 = 275,000.
4. **KNN Regression with k=3**: The three nearest neighbors are House 2 (1500 sq. ft.), House 3 (2000 sq. ft.), and House 4 (2500 sq. ft.). The predicted price is the average of their prices: (200,000 + 250,000 + 300,000) / 3 = 250,000.

Discussion:

This example shows how Linear Regression and KNN Regression handle predictions differently. Linear Regression gives a smooth, global prediction by fitting a single line across all data points, which works well when the relationship between features and the target is linear.
KNN Regression predictions, on the other hand, depend on local data points. With k=1, the prediction is more sensitive to local variations. As k increases, KNN averages more neighbors, resulting in smoother predictions. The choice of k influences how local or global the model's behavior becomes.

5. Suppose you have a dataset with three predictor variables X_1, X_2, and X_3, and a response variable Y. The data is given as follows:

Observation	**X_1**	**X_2**	**X_3**	**Y**
1	1	2	3	10
2	2	4	6	20

3	3	6	9	30
4	4	8	12	40

The goal is to predict Y using Principal Component Regression (PCR). Perform regression using the retained principal components and provide the regression equation.

Solution:

Step 1: Standardize the Variables

Calculate the mean and standard deviation for each variable X1, X2, and X3, and then standardize the variables.

| X̄1 = 2.5, σ_X1 = 1.118 |
| X̄2 = 5.0, σ_X2 = 2.236 |
| X̄3 = 7.5, σ_X3 = 3.354 |

Standardized Data:

X1*	X2*	X3*
-1.34	-1.34	-1.34
-0.45	-0.45	-0.45
0.45	0.45	0.45
1.34	1.34	1.34

Step 2: Compute Principal Components

Perform PCA on the standardized data and compute the principal components (PCs). The first principal component will be the linear combination of the original standardized variables that captures the most variance.

Assume the principal components are calculated as:

- PC1 = 0.577 × X1* + 0.577 × X2* + 0.577 × X3*

The principal component scores (projections) for the four data points would be:

| PC1 |

-2.32
-0.78
0.78
2.32

Step 3: Fit Regression Model

Now, regress Y on PC1:

$Y = \beta0 + \beta1 \times PC1$

Using linear regression, the coefficients could be computed, and suppose we get:

- $\beta0 = 25$
- $\beta1 = 6.45$

Thus, the final regression model is:

$Y = 25 + 6.45 \times PC1$

Step 4: Evaluate the Model

Finally, use this model to make predictions for Y and compare it with the actual values:

PC1	Predicted Y	Actual Y
-2.32	10	10
-0.78	20	20
0.78	30	30
2.32	40	40

In this case, the model perfectly predicts the dependent variable Y.