Assignment- 19th March

Q1. What is Min-Max Scaling, and how is it used in data preprocessing? Provide an example to illustrate its application.

Min-Max Scaling is a feature scaling technique that transforms data to a fixed range, usually [0, 1] or [-1, 1]. It preserves the relative relationships between data points.

Formula:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Example: Consider the dataset: [10, 20, 30, 40, 50]. To scale it to [0, 1]:

- $X_{\min} = 10, X_{\max} = 50$

Transformed values:

$$X' = \frac{X - 10}{50 - 10} = \frac{X - 10}{40}$$

Result: [0.0, 0.25, 0.5, 0.75, 1.0]

Q2. What is the Unit Vector technique in feature scaling, and how does it differ from Min-Max Scaling? Provide an example to illustrate its application.

Unit Vector Scaling scales each feature vector to have a unit norm (length = 1). It emphasizes the direction of data points rather than their magnitude.

Formula:

$$X' = \frac{X}{\|X\|}$$

Difference:

- Min-Max scaling maps features to a fixed range.
- Unit Vector scaling transforms each data point individually to have a unit length.

Example: For a vector [3, 4], the norm is:

$$\|X\| = \sqrt{3^2 + 4^2} = 5$$

Scaled vector:

$$X' = \left[\frac{3}{5}, \frac{4}{5}\right] = [0.6, 0.8]$$

Q3. What is PCA (Principal Component Analysis), and how is it used in dimensionality reduction? Provide an example to illustrate its application.

Principal Component Analysis (PCA) is a technique for reducing the dimensionality of data by transforming it into a set of orthogonal components called principal components, which capture the most variance in the data.

Steps:

1. Standardize the data.
2. Compute the covariance matrix.
3. Compute the eigenvectors and eigenvalues of the covariance matrix.
4. Project the data onto the principal components.

Example: If you have a dataset with 100 features, PCA can reduce it to a smaller number (e.g., 10) while retaining 95% of the variance.

Q4. What is the relationship between PCA and Feature Extraction, and how can PCA be used for Feature Extraction? Provide an example to illustrate this concept.

Relationship:

- PCA is a feature extraction technique that creates new features (principal components) as linear combinations of the original features.

Using PCA for Feature Extraction:

- PCA projects data onto fewer dimensions while preserving the variance.

Example: A dataset with features [height, weight, age] can be transformed into two principal components that explain most of the variance.

Q5. You are working on a project to build a recommendation system for a food delivery service. The dataset contains features such as price, rating, and delivery time. Explain how you would use Min-Max Scaling to preprocess the data.

Steps:

1. Identify the range of each feature (e.g., price, rating, delivery time).
2. Apply Min-Max scaling to each feature using:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

3. Ensure all features are scaled to a common range ([0, 1]) to prevent any single feature from dominating the recommendation model.

Q6. You are working on a project to build a model to predict stock prices. The dataset contains many features, such as company financial data and market

trends. Explain how you would use PCA to reduce the dimensionality of the dataset.

Steps:

1. **Standardize Data**: Scale all features to have a mean of 0 and a standard deviation of 1.
2. **Compute Covariance Matrix**: Calculate the relationships between features.
3. **Find Principal Components**: Identify the eigenvectors and eigenvalues of the covariance matrix.
4. **Retain Components**: Select components that explain a large percentage of the variance (e.g., 95%).
5. **Transform Data**: Project the dataset onto the selected principal components.

Q7. For a dataset containing the following values: [1, 5, 10, 15, 20], perform Min-Max scaling to transform the values to a range of -1 to 1.

Steps:

1. Find $X_{\text{min}} = 1$, $X_{\text{max}} = 20$.
2. Use the formula:

$$X' = 2 \times \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} - 1$$

3. Transform each value:
   - For $X = 1$:

$$X' = 2 \times \frac{1 - 1}{20 - 1} - 1 = -1$$

   - For $X = 20$:

$$X' = 2 \times \frac{20 - 1}{20 - 1} - 1 = 1$$

Result: Transformed values: [-1, -0.737, -0.263, 0.263, 1]

Q8. For a dataset containing the following features: [height, weight, age, gender, blood pressure], perform Feature Extraction using PCA. How many principal components would you choose to retain, and why?

Steps:

1. Standardize the features.

2. Compute the covariance matrix and eigenvalues.
3. Retain components that explain at least 95% of the variance.

Why: The number of principal components depends on the cumulative variance explained. If the first 2 components explain 95%, retain them.