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

Min-Max scaling (also known as normalization) rescales features to a specific range, typically between 0 and 1. It's calculated by subtracting the minimum value from each observation and then dividing by the range of the data.

Example: Suppose you have a feature with values ranging from 10 to 50. After applying Min-Max scaling, the values would be transformed to a range between 0 and 1, maintaining the proportion:

NewValue = OldValue − MinValue MaxValue − MinValue NewValue= MaxValue−MinValue OldValue−MinValue

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.

The Unit Vector technique scales each feature individually to have a unit norm (length of 1). It normalizes each sample (row) to have a unit norm, typically the Euclidean norm (L2 normalization).

Example: If you have a dataset where each sample (row) is represented as a vector, applying unit vector scaling will normalize each row vector to have a length of 1 while maintaining the direction of the vector.

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

PCA is a technique used for dimensionality reduction that identifies patterns and relationships in data by transforming it into a new coordinate system. It does this by finding the principal components, which are orthogonal directions in the feature space that capture the maximum variance in the data.

Example: In a dataset with multiple correlated features (e.g., height, weight, age), PCA can identify new uncorrelated variables (principal components) that retain most of the information. For instance, reducing these correlated features into two principal components that capture the most 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.

PCA can be used for Feature Extraction by transforming original features into a smaller set of principal components. These components are a linear combination of the original features and capture most of the data's variance.

Example: In image processing, consider a dataset of images represented by pixel values. Applying PCA can extract principal components that represent patterns (like edges,

textures) in the images, enabling dimensionality reduction while retaining most of the image information.

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.

For the food delivery dataset, Min-Max scaling can be applied to standardize the numerical features (price, rating, delivery time) within a specific range, say between 0 and 1. This ensures that all the features are on a comparable scale, preventing any particular feature from dominating the others due to differences in their magnitudes.

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.

For the stock price prediction dataset, PCA can be used to reduce the dimensionality by identifying the most important components that capture the variance in the financial and market trend features. By transforming these features into a smaller set of principal components, one can potentially reduce noise and multicollinearity while retaining essential information.

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.

To perform Min-Max scaling to a range of -1 to 1, you can use the formula:

NewValue = 2 × (OldValue − MinValue) MaxValue − MinValue − 1 NewValue= MaxValue−MinValue 2×(OldValue−MinValue)−1

Using this formula, with values ranging from 1 to 20:

MinValue = 1 , MaxValue = 20 MinValue=1,MaxValue=20

After applying the formula for each value, the Min-Max scaled values will be: − 1 , − 0.6 , − 0.2 , 0.2 , 0.6 −1,−0.6,−0.2,0.2,0.6

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?

The number of principal components to retain using PCA depends on the explained variance ratio or how much variance you want to retain in the data. You can analyze the cumulative explained variance to decide on the number of components to keep.

For instance, if you want to retain 95% of the variance, you can use PCA, calculate the explained variance for each component, and then choose the number of components that contribute to at least 95% of the total variance.

The decision of how many principal components to retain should balance the trade-off between dimensionality reduction and retaining enough information to maintain the model's predictive power.