**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 technique used to normalize the data within a specific range, typically between 0 and 1. It is especially useful when features have different units or scales. The formula for Min-Max scaling is:

Xscaled=X−XminXmax−XminX\_{\text{scaled}} = \frac{X - X\_{\text{min}}}{X\_{\text{max}} - X\_{\text{min}}}

Where:

* XscaledX\_{\text{scaled}} is the scaled value,
* XX is the original value,
* XminX\_{\text{min}} is the minimum value in the feature,
* XmaxX\_{\text{max}} is the maximum value in the feature.

**Example:** Given the dataset: [1, 5, 10, 15, 20], the Min-Max scaled values for a range of 0 to 1 would be calculated as follows:

* Minimum value: 1
* Maximum value: 20

Xscaled=X−120−1X\_{\text{scaled}} = \frac{X - 1}{20 - 1}

For X=1X = 1, Xscaled=1−119=0X\_{\text{scaled}} = \frac{1 - 1}{19} = 0.

For X=5X = 5, Xscaled=5−119=0.2105X\_{\text{scaled}} = \frac{5 - 1}{19} = 0.2105.

And so on, until all the values are scaled.

**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** (also known as normalization) scales the data so that each feature has a unit norm (i.e., the sum of squares of each feature's values equals 1). The formula used is:

Xnormalized=X∥X∥X\_{\text{normalized}} = \frac{X}{\|X\|}

Where ∥X∥\|X\| is the Euclidean norm (magnitude) of the vector.

**Difference from Min-Max scaling:**

* **Min-Max scaling** adjusts the data to fit within a specified range (typically 0 to 1).
* **Unit Vector Scaling** transforms each feature so that the feature vector has a length of 1, maintaining its direction while changing its magnitude.

**Example:** For a feature vector [1, 2, 3]:

* The Euclidean norm is ∥X∥=12+22+32=14\|X\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.
* The scaled vector would be [1,2,3]14\frac{[1, 2, 3]}{\sqrt{14}}.

Thus, each value in the vector is divided by 14\sqrt{14}, resulting in a normalized vector.

**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 statistical technique used to reduce the dimensionality of data while retaining as much variance as possible. PCA transforms the original features into new orthogonal features called principal components, ordered by the amount of variance they explain.

The steps involved in PCA:

1. Standardize the data.
2. Compute the covariance matrix of the data.
3. Calculate the eigenvalues and eigenvectors of the covariance matrix.
4. Sort the eigenvectors by eigenvalue in descending order.
5. Choose the top kk eigenvectors to form the new feature space.

**Example:** Imagine a dataset with two features: height and weight. PCA might reveal that the first principal component (PC1) explains most of the variance, while the second component (PC2) explains much less. By projecting the data onto the first principal component, you reduce the dataset's dimensionality from 2 to 1 while retaining most of the information.

**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 between PCA and Feature Extraction:**

* PCA can be used as a **feature extraction** technique by transforming the original features into new features (principal components) that are linear combinations of the original features.
* These new features (principal components) are typically uncorrelated and represent the most significant directions of variance in the data.

**Example:** For a dataset with features like height, weight, and age, PCA might combine these features into new components (e.g., one component representing overall body size). By selecting the top few components, you can extract the most significant features from the data.

**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.**

To preprocess the data using **Min-Max scaling**, the price, rating, and delivery time features should be scaled to a common range (typically between 0 and 1). This ensures that the features contribute equally to the recommendation model, preventing features with larger scales from dominating.

Steps:

1. Find the minimum and maximum values for each feature (price, rating, delivery time).
2. Apply Min-Max scaling to each feature.
3. The transformed features can then be used in the recommendation system model.

For example:

* Price: Min = $5, Max = $50, so price values are scaled to [0, 1].
* Rating: Min = 1, Max = 5, so rating values are scaled to [0, 1].
* Delivery Time: Min = 10 mins, Max = 60 mins, so delivery times are scaled to [0, 1].

**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.**

To reduce the dimensionality of a stock price prediction dataset using **PCA**:

1. Standardize the features (financial data, market trends) to have a mean of 0 and a standard deviation of 1.
2. Compute the covariance matrix of the standardized features.
3. Perform eigendecomposition to find the eigenvalues and eigenvectors.
4. Choose the top kk principal components that explain the most variance.
5. Project the original dataset onto these kk principal components, reducing the number of features.

This process helps focus on the most important features, improving model performance by reducing noise and complexity.

**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.**

The formula for Min-Max scaling to a range of [−1,1][-1, 1] is:

Xscaled=2×X−XminXmax−Xmin−1X\_{\text{scaled}} = 2 \times \frac{X - X\_{\text{min}}}{X\_{\text{max}} - X\_{\text{min}}} - 1

Where:

* Xmin=1X\_{\text{min}} = 1
* Xmax=20X\_{\text{max}} = 20

Let's apply this to the values:

* For X=1X = 1: 1−120−1=0⇒2×0−1=−1\frac{1 - 1}{20 - 1} = 0 \Rightarrow 2 \times 0 - 1 = -1
* For X=5X = 5: 5−119=0.2105⇒2×0.2105−1=−0.5789\frac{5 - 1}{19} = 0.2105 \Rightarrow 2 \times 0.2105 - 1 = -0.5789
* For X=10X = 10: 10−119=0.4737⇒2×0.4737−1=−0.0526\frac{10 - 1}{19} = 0.4737 \Rightarrow 2 \times 0.4737 - 1 = -0.0526
* For X=15X = 15: 15−119=0.7368⇒2×0.7368−1=0.4737\frac{15 - 1}{19} = 0.7368 \Rightarrow 2 \times 0.7368 - 1 = 0.4737
* For X=20X = 20: 20−119=1⇒2×1−1=1\frac{20 - 1}{19} = 1 \Rightarrow 2 \times 1 - 1 = 1

Thus, the scaled values are: [−1,−0.5789,−0.0526,0.4737,1][-1, -0.5789, -0.0526, 0.4737, 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?**

To determine how many principal components to retain:

1. Perform PCA and compute the explained variance ratio for each principal component.
2. Plot a cumulative explained variance curve to see how much variance is explained by the top components.
3. Retain enough components to explain, for example, 90-95% of the total variance.

If, for example, the first three principal components explain 95% of the variance, you would choose to retain those three components. This reduces dimensionality while retaining most of the data's information.