



Feature Engineering quiz

9 out of 9 correct

1. What is the purpose of Min-Max Scaling?

- ☐ To standardize data using mean and standard deviation
- ☒ To normalize data into a range of $[0,1]$
- ☐ To remove outliers from the data
- ☐ To perform feature extraction

Explanation: Min-Max Scaling, also known as normalization, transforms the data into a range of $[0,1]$ by subtracting the minimum value and dividing by the range. This is useful when the scale of the features varies widely, as it ensures that all features have equal importance in the analysis.

2. What is Feature Extraction?

- ☐ The process of selecting the most important features from a dataset
- ☒ The process of transforming data into a different representation
- ☐ The process of combining multiple datasets into a single dataset
- ☐ The process of removing outliers from a dataset

Explanation: Feature Extraction is the process of transforming data into a different representation that is more suitable for analysis. This can involve selecting the most important features, combining features, or creating new features based on the existing ones.



3. What is PCA (Principal Component Analysis)?

- ☒ A feature extraction technique that transforms data into a lower-dimensional space
- ☐ A classification algorithm used for image recognition
- ☐ A clustering algorithm used for unsupervised learning
- ☐ An outlier detection technique used for anomaly detection

Explanation: PCA (Principal Component Analysis) is a feature extraction technique that transforms data into a lower-dimensional space while preserving the most important information. It does this by finding the principal components, which are the linear combinations of the original features that explain the most variance in the data.

4. What is the purpose of PCA?

- ☐ To standardize data using mean and standard deviation
- ☐ To normalize data into a range of $[0,1]$
- ☒ To reduce the dimensionality of the data
- ☐ To remove outliers from the data

Explanation: The main purpose of PCA is to reduce the dimensionality of the data by finding the principal components that explain the most variance in the data. This is useful when dealing with high-dimensional datasets, as it allows for easier visualization and analysis of the data.

5. What is the range of values after performing Min-Max Scaling?

- ☐ $[-1, 1]$
- ☒ $[0, 1]$
- ☐ $[1, 10]$
- ☐ $[-10, 10]$

Explanation: Min-Max Scaling, also known as normalization, transforms the data into a range of $[0,1]$ by subtracting the minimum value and dividing by the range.

6. What is the L2 norm of a unit vector?

- ☒ 1
- ☐ 0
- ☐ -1
- ☐ Infinity

Explanation: A unit vector is a vector of length 1, and the L2 norm is the square root of the sum of the squared values. Therefore, the L2 norm of a unit vector is 1.

7. Which of the following is a feature extraction technique?

- ☐ Min-Max Scaling
- ☐ Unit Vector
- ☒ PCA
- ☐ K-Means Clustering

Explanation: PCA (Principal Component Analysis) is a feature extraction technique that transforms data into a lower-dimensional space while preserving the most important information.

8. A data scientist is working with a dataset of stock prices that vary widely in magnitude. Which of the following data preprocessing techniques should they use to ensure that all features have equal importance in the analysis?

- ☒ Min-Max Scaling
- ☐ Unit Vector
- ☐ Feature Extraction
- ☐ PCA

Explanation: Min-Max Scaling, also known as normalization, is useful when the scale of the features varies widely, as it ensures that all features have equal

importance in the analysis.

9. An analyst is working with a dataset of customer demographics and wants to reduce the number of features to improve the efficiency of the analysis. Which of the following techniques should they use?

- ☐ Min-Max Scaling
- ☐ Unit Vector
- ☐ Feature Extraction
- ☒ PCA

Explanation: PCA (Principal Component Analysis) is a feature extraction technique that transforms data into a lower-dimensional space while preserving the most important information. This is useful when dealing with high-dimensional datasets, as it allows for easier visualization and analysis of the data.

Submit