1. What exactly is a feature? Give an example to illustrate your point.

2. What are the various circumstances in which feature construction is required?

3. Describe how nominal variables are encoded.

4. Describe how numeric features are converted to categorical features.

5. Describe the feature selection wrapper approach. State the advantages and disadvantages of this approach?

6. When is a feature considered irrelevant? What can be said to quantify it?

7. When is a function considered redundant? What criteria are used to identify features that could be redundant?

8. What are the various distance measurements used to determine feature similarity?

9. State difference between Euclidean and Manhattan distances?

10. Distinguish between feature transformation and feature selection.

11. Make brief notes on any two of the following:

1. SVD (Standard Variable Diameter Diameter)
2. Collection of features using a hybrid approach
3. The width of the silhouette
4. Receiver operating characteristic curve

Answers:

1. In machine learning, a feature is a measurable attribute of the data that is used to make predictions or perform classification. For example, in an image classification task, the features could be the color, shape, and texture of the object in the image.
2. Feature construction is required in various circumstances, including:

* When the existing features are not informative enough to make accurate predictions or classification.
* When the data is in a raw format that needs to be processed and transformed into a usable format.
* When domain knowledge suggests that specific features should be included to improve the performance of the model.

1. Nominal variables are categorical variables that do not have a natural order. They are encoded using one-hot encoding, where each category is assigned a binary value (0 or 1) indicating whether it is present in the data sample. For example, if the nominal variable is "color" with categories "red," "green," and "blue," the encoding would be [1, 0, 0] for "red," [0, 1, 0] for "green," and [0, 0, 1] for "blue."
2. Numeric features can be converted to categorical features by discretizing them into a set of predefined bins. For example, if the numeric feature is "age," it can be converted to a categorical feature by dividing it into age groups (e.g., 0-18, 19-35, 36-50, 51+).
3. The feature selection wrapper approach involves training a model multiple times with different subsets of features and selecting the subset that results in the best performance. The advantage of this approach is that it considers the interaction between features and can identify subsets of features that work well together. However, it can be computationally expensive and may overfit to the specific dataset.
4. A feature is considered irrelevant if it does not contribute to the predictive power of the model. It can be quantified by measuring the decrease in model performance when the feature is removed from the dataset.
5. A function is considered redundant if it provides the same information as another feature. Criteria used to identify redundant features include measuring the correlation between features, testing whether the addition or removal of a feature improves the model's performance, and using techniques such as principal component analysis (PCA) to identify linear combinations of features.
6. The various distance measurements used to determine feature similarity include Euclidean distance, Manhattan distance, cosine similarity, Jaccard similarity, and Mahalanobis distance.
7. Euclidean distance is the straight-line distance between two points, while Manhattan distance is the sum of the absolute differences between the coordinates of the two points along each dimension. Euclidean distance is sensitive to outliers and assumes that all dimensions are equally important, while Manhattan distance is robust to outliers and considers each dimension independently.
8. Feature transformation involves applying mathematical functions to the existing features to create new features that better represent the data. Feature selection involves choosing a subset of the existing features that are most informative for the model.

I. SVD (Singular Value Decomposition) is a technique used to decompose a matrix into its constituent parts, allowing for the identification of the underlying structure in the data. It is often used in dimensionality reduction, where the high-dimensional data is transformed into a lower-dimensional space while preserving the most important information.

IV. Receiver operating characteristic (ROC) curve is a graphical representation of the performance of a binary classification model. It plots the true positive rate (TPR) against the false positive rate (FPR) at various classification thresholds, allowing for the evaluation of the tradeoff between sensitivity and specificity. A good model will have a curve that is closer to the upper left corner, indicating high TPR and low FPR