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

>>>A feature is an individual measurable property or characteristic of a phenomenon being observed or studied.

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

>>>When the raw data lacks relevant features.

When combining multiple sources of data.

To derive new insights from existing features.

When features need to be normalized or scaled.

3. Describe how nominal variables are encoded.

>>>Nominal variables are categorical variables without an inherent order.

Encoding methods include One-Hot Encoding and Label Encoding.

One-Hot Encoding: Converts each category into a binary column (0 or 1) to represent its presence.

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

>>>Numeric features can be converted to categorical by binning or discretization.

Binning: Divides the range of values into intervals (bins) and assigns a label to each interval.

Discretization: Converts continuous features into a discrete set of values, often represented by ranges or intervals.

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

>>>Involves selecting a subset of features based on model performance.

Steps: Iteratively add or remove features, train the model, and evaluate performance using cross-validation.

Advantages: Considers model performance directly, can find an optimal subset for a specific model.

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

>>>A feature is considered irrelevant when it doesn't contribute meaningful information to the task at hand.

Relevance can be quantified by measuring how much the feature influences the target variable or the model's performance.

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

>>>A function is redundant when its information is already captured by another function.

Criteria to identify redundancy include correlation between features, multicollinearity, and assessing their contribution to model performance.

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

>>>Euclidean Distance: Measures the straight-line distance between two points in Euclidean space.

Manhattan Distance: Measures the distance by summing the absolute differences between corresponding coordinates.

9. State difference between Euclidean and Manhattan distances?

>>>

10. Distinguish between feature transformation and feature selection.

>>>Feature Transformation: Involves creating new features using mathematical functions or transformations.

Feature Selection: Involves choosing a subset of existing features for model training, often based on their relevance and importance.

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

4. Receiver operating characteristic curveSVD (Singular Value Decomposition): A matrix factorization technique used in dimensionality reduction and data compression.

Collection of Features Using a Hybrid Approach: Combining various feature selection methods (filter, wrapper, embedded) for better results.

Silhouette Width: A measure of how close each instance in a cluster is to other instances in the same cluster compared to other clusters.

Receiver Operating Characteristic (ROC) Curve: Graphical representation of true positive rate against false positive rate to assess binary classifier performance.