Q1. What exactly is a feature? Give an example to illustrate your point.

* A feature is an individual attribute or property of data that can be used to make predictions or perform analysis.
* Example: In predicting house prices, features could include size, number of bedrooms, location, and age of the house.

Q2. What are the various circumstances in which feature construction is required?

* When data lacks relevant features.
* When creating new features that may improve predictive power.
* When converting complex data (e.g., text) to numerical form.
* When addressing data sparsity or dimensionality issues.

Q3. Describe how nominal variables are encoded.

* Nominal variables are categorical variables without inherent order.
* One-Hot Encoding: Convert each category into a binary column (0 or 1).
* Example: Encoding "Colors" variable with categories {Red, Green, Blue} into three binary columns.

Q4. Describe how numeric features are converted to categorical features.

* Binning or Discretization: Divide numeric values into ranges or bins and convert them into categorical labels.
* Example: Converting ages into categories like {Child, Teenager, Adult, Senior}.

Q5. Describe the feature selection wrapper approach. State the advantages and disadvantages of this approach?

* This approach selects features by repeatedly training models and evaluating their performance.
* Advantages: Considers model-specific performance, may lead to optimal feature subset.
* Disadvantages: Computationally expensive, may overfit training data.

Q6. 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 model's predictions.
* Quantification: Irrelevant features often have low correlation with the target variable.

Q7. When is a function considered redundant? What criteria are used to identify features that could be redundant?

* A function is redundant if its information is already captured by another feature.
* Criteria: High correlation between features may indicate redundancy.

Q8. What are the various distance measurements used to determine feature similarity?

Euclidean Distance, Manhattan Distance, Cosine Similarity, Jaccard Similarity are commonly used.

Q9. State difference between Euclidean and Manhattan distances?

* Euclidean Distance calculates the straight-line distance between two points in space.
* Manhattan Distance calculates the sum of the absolute differences between corresponding coordinates.

Q10. Distinguish between feature transformation and feature selection.

* Feature Transformation: Modifying existing features (e.g., scaling, PCA).
* Feature Selection: Selecting a subset of relevant features from the original set.

Q11. 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

* SVD (Singular Value Decomposition): Matrix factorization technique used for dimensionality reduction.
* Hybrid Feature Collection: Combination of various methods (e.g., filter, wrapper, embedded) for feature selection.
* Silhouette Width: Metric to measure the separation between clusters in unsupervised learning.
* Receiver Operating Characteristic (ROC) Curve: Graphical representation of a classification model's performance.