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

Solution - In machine learning, a feature is data that's used as the input for ML models to make predictions. Raw data is rarely in a format that is consumable by an ML model, so it needs to be transformed into features. This process is called feature engineering.

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

Solution - Feature construction, also known as feature engineering, refers to the process of creating new features or transforming existing features to improve the performance of machine learning models. There are several circumstances in which feature construction is required or beneficial:

Missing or Incomplete Data: If the dataset has missing values or incomplete data, feature construction techniques can be used to handle missing data, impute values, or create new features indicating the presence or absence of data.

Non-Linearity: In cases where the relationship between the features and the target variable is non-linear, feature construction techniques like polynomial features, interaction terms, or basis function expansion can help capture complex relationships and improve model performance**.**

Feature Scaling: Some machine learning algorithms are sensitive to the scale of the features. In such cases, feature construction techniques like standardization or normalization can be used to scale the features and bring them to a comparable range.

Categorical Data: Many machine learning algorithms require numerical input, so categorical variables need to be encoded into a numerical representation. Techniques like one-hot encoding, label encoding, or ordinal encoding can be used to transform categorical features into a suitable numerical format.

Dimensionality Reduction: High-dimensional datasets can suffer from the curse of dimensionality, leading to increased model complexity, computational inefficiency, and overfitting. Dimensionality reduction techniques like principal component analysis (PCA), linear discriminant analysis (LDA), or feature selection methods can be employed to reduce the number of features while preserving relevant information.

Noise Removal: If the dataset contains noisy or irrelevant features, feature construction techniques can be applied to identify and remove noisy features, reducing the impact of irrelevant information on the model.

3. **Describe how nominal variables are encoded.**

Solution - Nominal Encoding

When we have a feature where variables are just names and there is no order or rank to this variable's feature. For example: City of person lives in, Gender of person, Marital Status, etc… In the above example, We do not have any order or rank, or sequence.

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

Solution - Discretization: It is the process of transforming continuous variables into categorical variables by creating a set of intervals, which are contiguous, that span over the range of the variable's values. It is also known as “Binning”, where the bin is an analogous name for an interval.

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

Solution - In wrapper methods, the feature selection process is based on a specific machine learning algorithm that we are trying to fit on a given dataset. It follows a greedy search approach by evaluating all the possible combinations of features against the evaluation criterion.

The wrapper method has the advantages of better generalization and robust interaction with the classifier used for feature selection.

There are several disadvantages such as computationally intensive, discriminative power, lower shorter training times, classifier dependent selection, and higher risk of over-fitting than deterministic algorithms

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

Solution – One general definition for relevance is that a feature can be regarded as irrelevant if it is conditionally independent of the class labels or it does not influence the class labels; in these cases, it can be discarded.

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

Solution - Redundant Features Slow Down the Training Process. The more features you have, the slower the calculations are. However, there is another hidden factor that slows down training significantly.

if two features {X1, X2} are highly correlated, then the two features become redundant features since they have same information in terms of correlation measure. In other words, the correlation measure provides statistical association between any given a pair of features.

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

Solution – Some of the methods for similarity measures between two data points include Euclidean distance, Manhattan distance, Minkowski distance, and Chebyshev distance.

9. **State difference between Euclidean and Manhattan distances?**

Solution - Euclidean distance is the shortest path between source and destination which is a straight line but Manhattan distance is sum of all the real distances between source(s) and destination(d) and each distance are always the straight lines.

10. **Distinguish between feature transformation and feature selection.**

Solution - feature transformation: transformation of data to improve the accuracy of the algorithm; feature selection: removing unnecessary features.

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

1.SVD (Standard Variable Diameter Diameter)

Solution - SVD is basically a matrix factorization technique, which decomposes any matrix into 3 generic and familiar matrices.

2. Collection of features using a hybrid approach

Solution - A hybrid feature selection method is proposed for classification in small sample size data sets. The filter step is based on instance learning taking advantage of the small sample size of data.

3. The width of the silhouette

Solution - In this work, we use the Silhouette index to evaluate the clustering performance. The Silhouette width s x i for the point x i is defined as [11]: s x i = b x i − a x i max b x i , a x i . b x i = min d l x i , among all clusters l ≠ k

4. Receiver operating characteristic curve

Solution - An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate False Positive Rate