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

A “feature” is any measurable input that can be used in a predictive model — it could be the color of an object or the sound of someone's voice.

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

can add more information and give more insights of the data we are dealing with.

deriving features from the pre-existing features and coming up with more meaningful features

3. Describe how nominal variables are encoded.

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.

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

Many machine-learning algorithms, such as naive Bayes classification and prediction, work only with categorical data. So if your raw data is numeric and you want to apply naive Bayes, you have to discretize the data. You might also have mixed numeric and categorical data, such as the data often found in an Excel spreadsheet. Very few machine-learning algorithms work with mixed data, so you can convert the numeric data to categorical data and then use a machine-learning algorithm that works with categorical data.

For example, if you have a data set of people’s heights in inches, such as 59.5, 64.0 and 75.5, you might want to convert this numeric data into categorical data, for example 0, 1, and 2, to represent short, medium, and tall.

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

The wrapper classification algorithms with joint dimensionality reduction and classification can also be used but these methods have high computation cost, lower discriminative power. Moreover, these methods depend on the efficient selection of classifiers for obtaining high accuracy.

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

Irrelevant or partially relevant features can negatively impact model performance. Feature Selection is one of the core concepts in machine learning which hugely impacts the performance of your model. The data features that you use to train your machine learning models have a huge influence on the performance you can achieve.

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

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?

The similarity measure is a distance with dimensions describing object features. That means if the distance among two data points is small then there is a high degree of similarity among the objects and vice versa. The similarity is subjective and depends heavily on the context and application.

9. State difference between Euclidean and Manhattan distances?

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.

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