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

A feature is one column of the data in your input set. For instance, if you're trying to predict the type of pet someone will choose, your input features might include age, home region, family income, etc. The label is the final choice, such as dog, fish, iguana, rock, etc .

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

Feature Construction is a useful process as it can add more information and give more insights of the data we are dealing with.

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

Nominal variables are encoded by using One-Hot Encoding.

It maps each category with binary numbers (0 or 1). This type of encoding is used when the data is nominal. Newly created binary features can be considered dummy variables. After one hot encoding, the number of dummy variables depends on the number of categories presented in the data

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

There are many ways in which conversion can be done , one such way is by using Pandas' integrated cut-function. Pandas' cut function is a distinguished way of converting numerical continuous data into categorical data.

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

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.

This methods have high computation cost, lower discriminative power.

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

* 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.

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

A function is considered redundant when there is any disruption of system operation in the case of a technical failure or disaster by maintaining a continuity of service.

Redundant features are those that are correlated with other features and not relevant in the sense that they do not improve the discriminatory ability of a set of features.

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

The most commonly used distance measures in machine learning are as follows: Hamming Distance. Euclidean Distance. Manhattan Distance

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.**

The main difference : Feature Extraction transforms an arbitrary data, such as text or images, into numerical features that is understood by machine learning algorithms. Feature Selection on the other hand is a machine learning technique applied on these (numerical) features

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

1.SVD (Standard Variable Diameter Diameter)

SVD allows us to extract and untangle information.

2. **Collection of features using a hybrid approach**

A hybrid feature selection method is proposed for classification in small sample size data sets.

**3. The width of the silhouette**

The Average Silhouette Width (ASW) of a clustering is the average distance of to points in the cluster to which it was assigned, and is the average distance of to the points in the nearest cluster to which it was not assigned.

**4. Receiver operating characteristic curve**

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds.