1)What is the KNN algorithm?

Ans- The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

2) How do you choose the value of K in KNN?

Ans- The choice of k will largely depend on the input data as data with more outliers or noise will likely perform better with higher values of k. Overall, it is recommended to have an odd number for k to avoid ties in classification, and cross-validation tactics can help you choose the optimal k for your dataset.

3) What is the difference between KNN classifier and KNN regressor?

Ans- The key differences are: KNN regression tries to predict the value of the output variable by using a local average. KNN classification attempts to predict the class to which the output variable belong by computing the local probability.

4) How do you measure the performance of KNN?

Ans- The main concept for k-NN depends on calculating the distances between the tested, and the training data samples in order to identify its nearest neighbours. The tested sample is then simply assigned to the class of its nearest neighbor

5) What is the curse of dimensionality in KNN?

Ans- The “Curse of Dimensionality” is a tongue in cheek way of stating that there's a ton of space in high-dimensional data sets. The size of the data space grows exponentially with the number of dimensions. This means that the size of your data set must also grow exponentially in order to keep the same density.

6) How do you handle missing values in KNN?

Ans- The idea in kNN methods is to identify 'k' samples in the dataset that are similar or close in the space. Then we use these 'k' samples to estimate the value of the missing data points. Each sample's missing values are imputed using the mean value of the 'k'-neighbors found in the dataset.

7) Compare and contrast the performance of the KNN classifier and regressor. Which one is better for which type of problem?

Ans- The key differences are: KNN regression tries to predict the value of the output variable by using a local average. KNN classification attempts to predict the class to which the output variable belong by computing the local probability.

KNN is better than linear regression when the data have high SNR

The most significant difference between regression vs classification is that while regression helps predict a continuous quantity, classification predicts discrete class labels. There are also some overlaps between the two types of machine learning algorithms.

8) What are the strengths and weaknesses of the KNN algorithm for classification and regression tasks, and how can these be addressed?

Ans-

It's easy to understand and simple to implement.

It can be used for both classification and regression problems.

It's ideal for non-linear data since there's no assumption about underlying data.

It can naturally handle multi-class cases.

It can perform well with enough representative data.

It has advantages - nonparametric architecture, simple and powerful, requires no traning time, but it also has disadvantage - memory intensive, classification and estimation are slow

9) What is the difference between Euclidean distance and Manhattan distance in KNN?

Ans- Euclidean distance is the shortest path between source and destination which is a straight line as shown in Figure 1.3. but Manhattan distance is sum of all the real distances between source(s) and destination(d) and each distance are always the straight lines

We don't use Manhattan Distance, because it calculates distance horizontally or vertically only. It has dimension restrictions. On the other hand, the Euclidean metric can be used in any space to calculate distance. Since the data points can be represented in any dimension, it is a more viable option.

10) What is the role of feature scaling in KNN?

Ans- Feature scaling is required to get the better performance of the KNN algorithm. For Example, Imagine a dataset having n number of instances and N number of features. There is one feature having values ranging between 0 and 1. Meanwhile, there is also a feature that varies from -999 to 999.