1. In the sense of machine learning, what is a model? What is the best way to train a model?

A: A machine learning model is like an instance of a machine learning algorithm. Each algorithm has some coefficients that need to be found. This finding of coefficients is done during model training. Models can be trained by applying algorithms on data.

2. In the sense of machine learning, explain the "No Free Lunch" theorem.

A: The average performance across all possible problems is the same for all optimization algorithms. Meaning that no algorithm works for all the possible problems and we have to choose the algorithm that works best for the scenario.

3. Describe the K-fold cross-validation mechanism in detail.

A: In K-fold CV, the training data set is divided into K smaller datasets(folds). Out of which K-1 folds are used as training data sets and the remaining fold is used to validate the model. This process continues for K times with validation set changing each time. Performance of model is evaluated after each iteration and at the end of all the iterations, it is checked whether the model is performing in a similar manner across the iterations.

4. Describe the bootstrap sampling method. What is the aim of it?

A: It is used to estimate statistics of a larger dataset by using statistics from a group of smaller datasets that are randomly picked with replacement from the original data set.

5. What is the significance of calculating the Kappa value for a classification model? Demonstrate how to measure the Kappa value of a classification model using a sample collection of results.

A: Kappa is an evaluation metric used for highly imbalanced data sets. It is calculated by using the formula $(p_o - p_e)/(1-p_e)$

6. Describe the model ensemble method. In machine learning, what part does it play?

A: An ensemble model combines multiple simpler models to make predictions. Ensembles help in increasing accuracy of the predictions without overfitting the data.

7. What is a descriptive model's main purpose? Give examples of real-world problems that descriptive models were used to solve.

A: The purpose of a descriptive model is to be interpretable to humans, i.e, humans can understand the logic behind predictions. Clustering is a kind of descriptive model that can be used to solve user segmentation, image processing and so on

8. Describe how to evaluate a linear regression model.

A: There are a few metrics that can be used to evaluate a linear regression model. Most widely used metric is R value followed by metrics such as RMSE, F-statistic, p values and so on. Linear regression models can also be evaluated based on assumptions of a linear regression.

9. Distinguish :

1. Descriptive vs. predictive models

A: Descriptive models are focused on interpretability while predictive models focus on predicting power.

2. Underfitting vs. overfitting the model

A: When a model fails to capture some trends in the data, it is called underfitting. On the other hand, if a model captures errors as a trend, it is called overfitting.

3. Bootstrapping vs. cross-validation

A: Bootstrapping relies on random sampling while cross validation does not.

10. Make quick notes on:

1. LOOCV.

A: Leave One Out Cross Validation: Leaves one observation in the training set as a validation set.

2. F-measurement

A: F1 score of F measure is a combination of precision and recall.

3. The width of the silhouette

A: It is a measure of closeness of a point to its cluster compared to the nearest other cluster.

4. Receiver operating characteristic curve

A: A curve between true positive rate and false positive rate. This curve helps in evaluating a classification model.