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

A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output.

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

The theorem states that all optimization algorithms perform equally well when their performance is averaged across all possible problems

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

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into

The general procedure is as follows:

1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
   1. Take the group as a hold out or test data set
   2. Take the remaining groups as a training data set
   3. Fit a model on the training set and evaluate it on the test set
   4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores

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

Bootstrapping assigns measures of accuracy (bias, variance, confidence intervals, prediction error, etc.) to sample estimates. This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods.

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.

It basically tells you how much better your classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class. Cohen's kappa is always less than or equal to 1. Values of 0 or less, indicate that the classifier is useless.

The kappa statistic is used to control only those instances that may have been correctly classified by chance. This can be calculated using both the observed (total) accuracy and the random accuracy. Kappa can be calculated as: Kappa = (total accuracy – random accuracy) / (1- random accuracy).

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

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

Ensemble learning is the process by which multiple models, such as classifiers or experts, are strategically generated and combined to solve a particular computational intelligence problem. Ensemble learning is primarily used to improve the (classification, prediction, function approximation, etc.)

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

A descriptive model describes a system or other entity and its relationship to its environment. It is generally used to help specify and/or understand what the system is, what it does, and how it does it. A geometric model or spatial model is a descriptive model that represents geometric and/or spatial relationships.

Descriptive modeling is a mathematical process that describes real-world events and the relationships between factors responsible for them. The process is used by consumer-driven organizations to help them target their marketing and advertising efforts.

8. Describe how to evaluate a linear regression model.

There are 3 main metrics for model evaluation in regression:

1. R Square/Adjusted R Square.
2. Mean Square Error(MSE)/Root Mean Square Error(RMSE)
3. Mean Absolute Error(MAE)

9. Distinguish :

1. Descriptive vs. predictive models

A descriptive model will exploit the past data that are stored in databases and provide you with the accurate report. In a Predictive model, it identifies patterns found in past and transactional data to find risks and future outcomes.

2. Underfitting vs. overfitting the model

Your model is underfitting the training data when the model performs poorly on the training data. Your model is overfitting your training data when you see that the model performs well on the training data but does not perform well on the evaluation data.

3. Bootstrapping vs. cross-validation

Cross validation splits the available dataset to create multiple datasets, and Bootstrapping method uses the original dataset to create multiple datasets after resampling with replacement

10. Make quick notes on:

1. LOOCV.

LOOCV(Leave One Out Cross-Validation) is a type of cross-validation approach in which each observation is considered as the validation set and the rest (N-1) observations are considered as the training set. In LOOCV, fitting of the model is done and predicting using one observation validation set.

In this approach, we reserve only one data point from the available dataset, and train the model on the rest of the data. This process iterates for each data point. This also has its own advantages and disadvantages. Let’s look at them:

* We make use of all data points, hence the bias will be low
* We repeat the cross validation process n times (where n is number of data points) which results in a higher execution time
* This approach leads to higher variation in testing model effectiveness because we test against one data point. So, our estimation gets highly influenced by the data point. If the data point turns out to be an outlier, it can lead to a higher variation

2. F-measurement

The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

When Precision and Recall both are the important feature at that time F measurement score are important features to used.

3. The width of the silhouette

The Average Silhouette Width (ASW) of a clustering is ̄ a ( i ) 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

The ROC curve shows the trade-off between sensitivity (or TPR) and specificity (1 – FPR). Classifiers that give curves closer to the top-left corner indicate a better performance. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.