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

Follow basic steps:

1. Data gathering

2. Data Pre-processing

3. Exploratory Data Analysis

4. Model selection

5. Training Machine Learning Models

Regression

Classification

Clustering

6. Testing

7. Monitoring

8. Model Retraining

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.

It implies that there is no single best optimization algorithm. Because of the close relationship between optimization, search, and machine learning, it also implies that there is no single best machine learning algorithm for predictive modeling problems such as classification and regression.

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

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

Ans- Sampling with replacement . It is used in bagging techiniques(to sample data in different groups and then feed them in different models and compare output. The most frequent result wins)

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.

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.

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. To better understand this definition lets take a step back into ultimate goal of machine learning and model building

7. What is a descriptive model’s main purpose? Give examples of real-world problems that

descriptive models were used to solve.

Descriptive models main purpose is to summarize a dataset via central tendency or measure of spread(dispersion)

8. Describe how to evaluate a linear regression model.

Accuracy score, r2 statistics ,modified r2 statistics

9. 9. Distinguish :

1. Descriptive vs. predictive models
2. Desciptive model helps to understand data by correlation etc whereas predictive is used to predict output from that model.
3. Underfitting vs. overfitting the model – underfitting(model has poor accuracy on both training and testing dataset)

Overfitting(model has best accuracy on train whereas worst on test dataset.

3. Bootstrapping vs. cross-validation

Bootstrapping is sampling with replacement.

Cross validation simply sampling into subsets