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

Ans: **the machine learning model takes the inputs and predicts the output. The best to train a model is to split into train and test split and also use cross validation to further split the train test into k parts so all the training data set is used for both training and testing.**

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

Ans: **The “no free lunch” (NFL) theorem for supervised machine learning is a theorem that essentially implies that no single machine learning algorithm is universally the best-performing algorithm for all problems**

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

Ans: **K fold cross validation in implemented by cross\_val\_score method or GridSearchCV method in machine learning which allows you to split your training data into k fold where in each iteration a different k-1 subset is used for training and the remaining subset is for testing. That way all of the training data points are used for training and testing in different subsets of the training data. Which, in result, gives you more accurate estimate and it validates the performance of the model on multiple folds of your data.**

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

Ans: **The bootstrap method is a statistical technique for estimating quantities about a population by averaging estimates from multiple small data samples. Importantly, samples are constructed by drawing observations from a large data sample one at a time and returning them to the data sample after they have been chosen. This process allows for the calculation of standard errors, confidence intervals, and hypothesis testing**

1. 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.

Ans: **Cohen's kappa is a metric often used to assess the agreement between two raters. It can also be used to assess the performance of a classification model. Below is the formula for cohen kappa’s formula:**

Text, table, letter

Description automatically generated

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

Ans: **Ensemble technique is used for combining homogeneous or heterogenous algorithms in machine learning to get a better model. Bagging and Boosting are Homogeneous ensemble techniques where same type of multiple models are used. Bagging is considered a parallel approach and Boosting is a sequential approach. Stacking is an example of Heterogeneous ensemble method.**

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

Ans: **descriptive models are used to create relationship in the data like finding the central tendency of the data or grouping the data and find its relationship with the existing data. This model is used to describe the existing or historical data to infer some knowledge out of that. For example, a sale company can use the previous year sale data to see which month they sold the items the most and they can plan accordingly for the same month.**

8. Describe how to evaluate a linear regression model.

Ans: **different matrixes can be employed to evaluate a linear regression among which R square and Adjusted R2 are used predominantly. R square is used when there are only two variables in a data set but not suitable for more variables in a data set as R square value tends to increase with every extra feature demonstrating incorrect accuracy results irrespective of the correlation of variables. To solve this problem, Adjusted R square is used where the value of adjusted R square only increases if there is a correlation between newly added independent and dependent variable showing accurate results.**

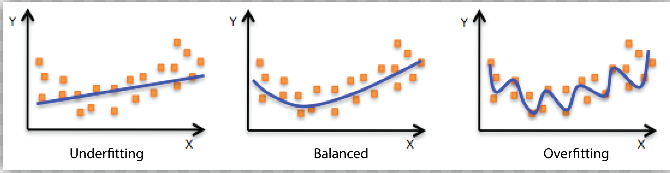
9. Distinguish :

1. Descriptive vs. predictive models:

Ans: **descriptive modelling as name suggest it is used to delve into historical data to see what has happened in the past. It describes the data which we already have. For examples, creating a sales report to see who sales took place last month is an example of descriptive modelling. On the contrary, predictive modelling is used to predict the outcome based on the historical data. We use several ML algorithms for forecasting. For example, predictive the sales figure for next month is an example of predictive modelling.**

2. Underfitting vs. overfitting the model:

**Ans: A model is considered overfitted when the best fit line tries to pass through all the data points in the data set. It is undesirable in ML as it doesn’t behave well with the test data or unseen data therefore, several techniques such Ridge algorithm is used to eliminate overfitting. Underfitting is the case when the model doesn’t work well even with the training data and gives poor results related to the training data. It is shown as below:**



3. Bootstrapping vs. cross-validation: **bootstrapping is a random sampling where same elements can be repeated multiple times in the split but cross validation is not a random sampling and the same element can’t be repeated in multiple splits.**

10. Make quick notes on:

1. LOOCV: **It is called leave one out which is special type of K fold cross validation but it could be time consuming due to large value of N.**

2. F-measurement: **F-Beta score is calculated when FP and FN both are important. When both are equally important then beta equals to 1.**

3. The width of the silhouette: **this is used to identify if an element belongs to its own cluster more than the neighbouring cluster.**

4. Receiver operating characteristic curve:: **ROC is used to measure the performance of a classification problem and decide the threshold value.**