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

A: Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples. In supervised learning, a machine learning algorithm builds a model by examining many examples and attempting to find a model that minimizes loss; this process is called empirical risk minimization.

* Reframe the problem. Sometimes, improving a model may have nothing to do with the data or techniques used to train the model
* Provide more data samples
* Add context to the data
* Use meaningful data and features
* Cross-validation
* Hyperparameter tuning
* Choose a different algorithm

1. In the sense of machine learning, explain the &quot;No Free Lunch&quot; theorem.

A:

The “*No Free Lunch*” Theorem argues that, without having substantive information about the modeling problem, there is no single model that will always do better than any other model. Because of this, a strong case can be made to try a wide variety of techniques, then determine which model to focus on.

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

A:

K-fold Cross-Validation is when the dataset is split into a K number of folds and is used to evaluate the model's ability when given new data. K refers to the number of groups the data sample is split into. For example, if you see that the k-value is 5, we can call this a 5-fold cross-validation.

Mechanism

* Split the entire data randomly into K folds (value of K shouldn’t be too small or too high, ideally we choose 5 to 10 depending on the data size). The higher value of K leads to less biased model (but large variance might lead to over-fit), where as the lower value of K is similar to the train-test split approach we saw before.
* Then fit the model using the K-1 (K minus 1) folds and validate the model using the remaining Kth fold. Note down the scores/errors.
* Repeat this process until every K-fold serve as the test set. Then take the average of your recorded scores. That will be the performance metric for the model.

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

Bootstrap sampling is used in a machine learning ensemble algorithm called bootstrap aggregating (also called bagging). It helps in avoiding [overfitting](https://www.analyticsvidhya.com/blog/2020/02/underfitting-overfitting-best-fitting-machine-learning/) and improves the stability of [machine learning algorithms](https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/).

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:

The kappa score is an interesting metric. Its origins are in the field of psychology: it is used for measuring the agreement between two human evaluators or raters (e.g., psychologists) when rating subjects (patients). It was later “appropriated” by the machine-learning community to measure classification performance

To calculate the Kappa coefficient we will take the probability of agreement minus the probability of disagreement divided by 1 minus the probability of disagreement. This is a positive value which means there is some mutual agreement between the parties.

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

A:

The ensemble methods in machine learning combine the insights obtained from multiple learning models to facilitate accurate and improved decisions. These methods follow the same principle as the example of buying an air-conditioner cited above.

In learning models, noise, variance, and bias are the major sources of error. The ensemble methods in machine learning help minimize these error-causing factors, thereby ensuring the accuracy and stability of machine learning (ML) algorithms.

7. What is a descriptive model&#39;s main purpose? Give examples of real-world problems that

descriptive models were used to solve.

A:

A descriptive model is used for tasks that would benefit from the insight gained from summarizing data in new and interesting ways. As opposed to predictive models that predict a target of interest, in a descriptive model, no single feature is more important than any other.

Summarising past events such as sales and operations data or marketing campaigns. Social media usage and engagement data such as Instagram or Facebook likes. Reporting general trends. Collating survey results.

8. Describe how to evaluate a linear regression model.

A:

There are 3 main metrics for model evaluation in regression:

* R Square/Adjusted R Square.
* Mean Square Error(MSE)/Root Mean Square Error(RMSE)
* Mean Absolute Error(MAE)

9. Distinguish :

1. Descriptive vs. predictive models

A:

Descriptive data mining is a data mining technique that analyzes the past data to provide latest information on past events, while predictive data mining is a data mining technique that is used to analyze past data and provides answers of future queries

2. Underfitting vs. overfitting the model

A:

Under-observing the features leads to a higher error in the training and unseen data samples. It is different from overfitting, where the model performs well in the training set but fails to generalize the learning to the testing set.

3. Bootstrapping vs. cross-validation

A:

Bootstrapping contains repeated elements in every subset. Bootstrapping relies on random sampling. Cross-validation does not rely on random sampling, just splitting the dataset into k unique subsets. Cross-validation is usually used to test an ML model's generalization capabilities.

10. Make quick notes on:

1. LOOCV.

A:

The Leave-One-Out Cross-Validation, or LOOCV, procedure is used to estimate the performance of machine learning algorithms when they are used to make predictions on data not used to train the model.

2. F-measurement

A:

In [statistical](https://en.wikipedia.org/wiki/Statistics) analysis of [binary classification](https://en.wikipedia.org/wiki/Binary_classification), the F-score or F-measure is a measure of a test's [accuracy](https://en.wikipedia.org/wiki/Accuracy_and_precision#In_binary_classification). It is calculated from the [precision](https://en.wikipedia.org/wiki/Precision_(information_retrieval)) and [recall](https://en.wikipedia.org/wiki/Recall_(information_retrieval)) of the test, where the precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive. Precision is also known as [positive predictive value](https://en.wikipedia.org/wiki/Positive_predictive_value), and recall is also known as [sensitivity](https://en.wikipedia.org/wiki/Sensitivity_and_specificity) in diagnostic binary classification.

3. The width of the silhouette

A:

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.

1. Receiver operating characteristic curve

A:

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters: True Positive Rate. False Positive Rate.