Daily Machine Learning Interview Questions

71. Differentiate between Statistical Modeling and Machine Learning?

Machine learning models are about making accurate predictions about the situations, like Foot Fall in restaurants, Stock-Price, etc. where-as, Statistical models are designed for inference about the relationships between variables, as What drives the sales in a restaurant, is it food or Ambience

72. Differentiate between Boosting and Bagging?

Bagging and Boosting are variants of Ensemble Techniques. Bootstrap Aggregation or bagging is a method that is used to reduce the variance for algorithms having very high variance. Decision trees are a particular family of classifiers which are susceptible to having high bias.

Decision trees have a lot of sensitiveness to the type of data they are trained on. Hence generalization of results is often much more complex to achieve in them despite very high fine-tuning. The results vary greatly if the training data is changed in decision trees.

Hence bagging is utilised where multiple decision trees are made which are trained on samples of the original data and the final result is the average of all these individual models.

Boosting is the process of using an n-weak classifier system for prediction such that every weak classifier compensates for the weaknesses of its classifiers. By weak classifier, we imply a classifier which performs poorly on a given data set.

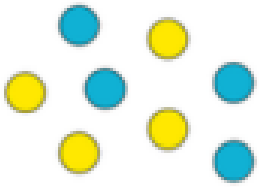
It’s evident that boosting is not an algorithm rather it’s a process. Weak classifiers used are generally logistic regression, shallow decision trees etc. There are many algorithms which make use of boosting processes but two of them are mainly used: Adaboost and Gradient Boosting and XGBoost.

73 . What are hyperparameters and how are they different from parameters?

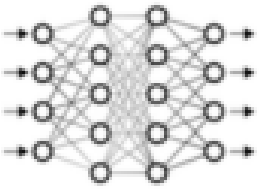
A parameter is a variable that is internal to the model and whose value is estimated from the training data. They are often saved as part of the learned model. Examples include weights, biases etc.

A hyperparameter is a variable that is external to the model whose value cannot be estimated from the data. They are often used to estimate model parameters. The choice of parameters is sensitive to implementation. Examples include learning rate, hidden layers etc

Parameters vs. Hyperparameters



Training data



w1 = 0.213
w2 = 1.645

Parameters

Number of layers = 3
Learning rate = 0.01

Hyperparameters

74. How to deal with multicollinearity?

Multi collinearity can be dealt with by the following steps:

- Remove highly correlated predictors from the model.
- Use Partial Least Squares Regression (PLS) or Principal Components Analysis.

75. What is Heteroscedasticity?

It is a situation in which the variance of a variable is unequal across the range of values of the predictor variable.

It should be avoided in regression as it introduces unnecessary variance.

Thank You