INSAID

DATA SCIENCE INTERVIEW QUESTIONS

Part-3

@insaid.learn

INSAID

DEFINE BIAS-VARIANCE TRADE-OFF

LET US FIRST UNDERSTAND THE MEANING OF BIAS & VARIANCE IN DETAIL:

BIAS

- It is a kind of error in a machine learning model when an ML Algorithm is oversimplified. When a model is trained, at that time it makes simplified assumptions so that it can easily understand the target function.

- Some algorithms that have low bias are Decision Trees, SVM, etc.

- Logistic and Linear Regression algorithms are the ones with a high bias.

Swipe!

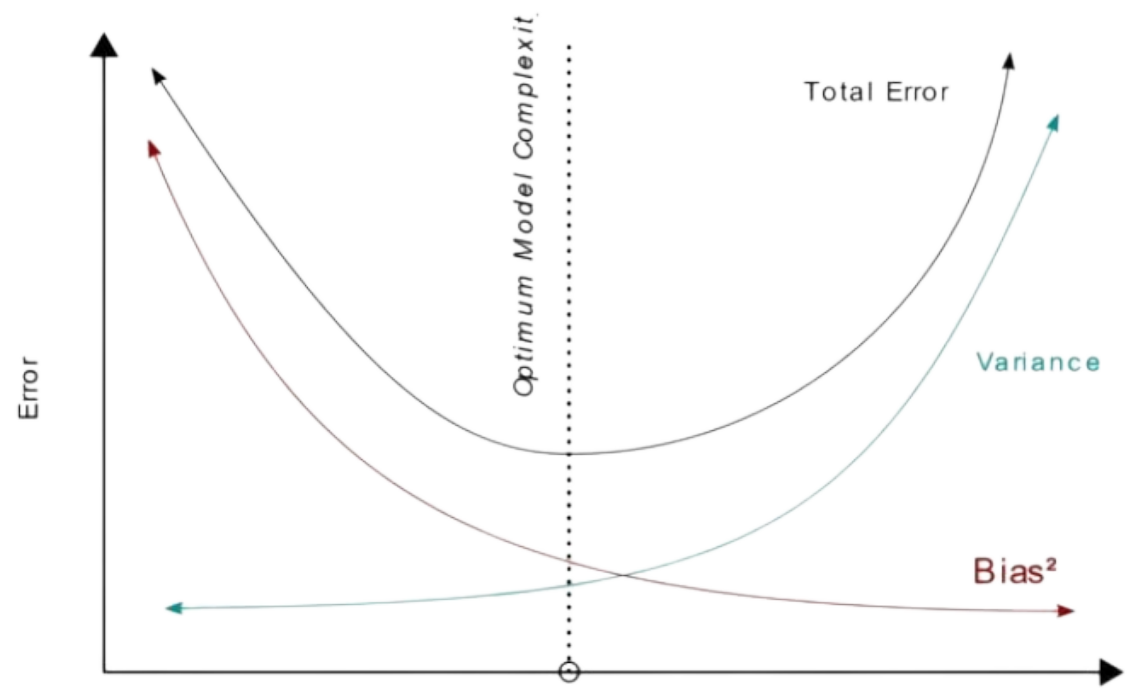
VARIANCE

- It is also a kind of an error. It's introduced into an ML model when an ML algorithm is made highly complex. This model also learns noise from the data set that is meant for training. It further performs badly on the test data set. This may lead to over lifting as well as high sensitivity.

- When the complexity of a model is increased, a reduction in the error is seen. This is caused by the lower bias in the model. But, this does not happen till we reach the optimal point. After this point, if we keep on increasing the complexity of the model, it will be over lifted and will suffer from the problem of high variance.

Swipe!

LET'S REPRESENT THIS SITUATION WITH THE HELP OF A GRAPH:



- As you can see from the image above, before the optimal point, increasing the complexity of the model reduces the error (bias). However, after the optimal point, we see that the increase in the complexity of the machine learning model increases the variance. MSE = Bias^2 + Variance

@insaid.learn

Swipe!

TRADE-OFF OF BIAS & VARIANCE

- So, as we know that bias and variance, both are errors in machine learning models. It is very essential that any machine learning model has low variance as well as low bias so that it can achieve good performance.
- So the trade-off is simple. If we increase th bias, the variance will decrease and vice versa.

INSAID

WHAT IS LOGISTIC REGRESSION?

@insaid.learn

LOGISTIC REGRESSION

- It is also known as the logit model. Its is a technique to predict outcome from a linear combination of variables (called the predictor variables).

- For example, let us say that we want to predict the outcome of elections for a particular political leader. So, we want to find out whether this leader is going to win the election or not. So, the result is binary i.e. win (1) or loss (0). However, the input is a combination of linear variables like the money spent on advertising, the past work done by the leader and the party, etc.

Swipe!

INSAID

WHAT IS LINEAR REGRESSION?

@insaid.learn

LINEAR REGRESSION

- **It is a technique in which the score of a variable Y is predicted using the score of a predictor variable X. Y is called the criterion variable. Some of the drawbacks of Linear Regression are as follows:**

- The assumption of linearity of errors is a major drawback.

- It cannot be used for binary outcomes. We have Logistic Regression for that.

- Overfitting problems are there that can't be solved.

@insaid.learn

Swipe!

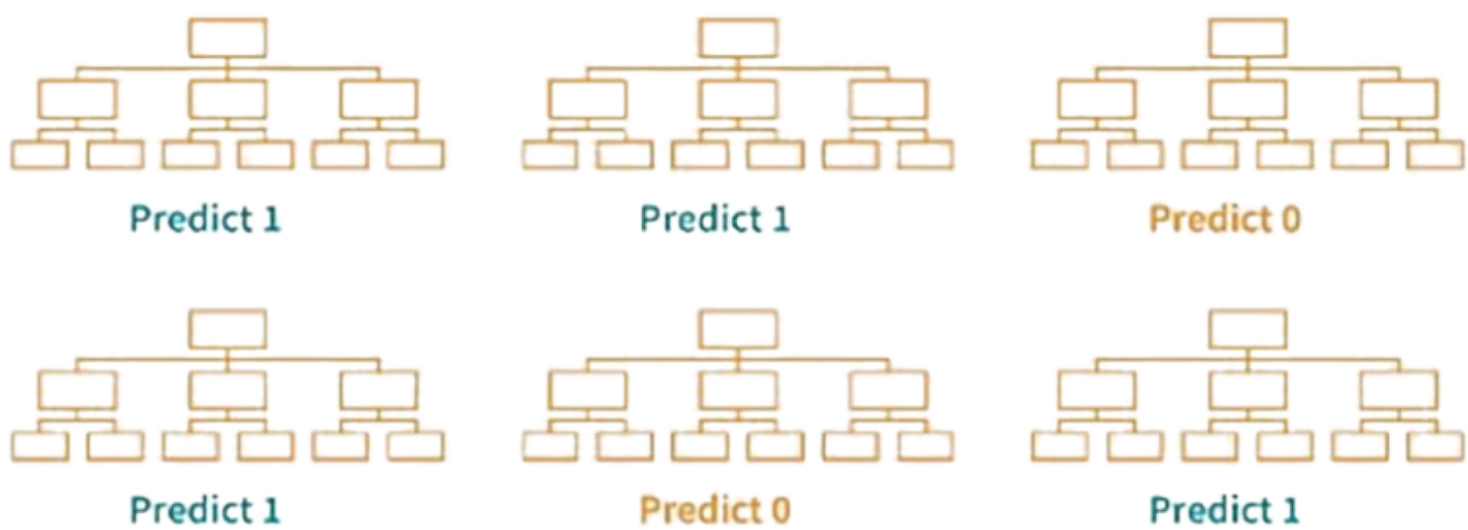
INSAID

WHAT IS RANDOM FOREST?

@insaid.learn

RANDOM FOREST

- It consists of a large number of decision trees that operate as an ensemble. Basically, each tree in the forest gives a class prediction and the one with the maximum number of votes comes the prediction of our model.
- For instance, in the example shown below, 4 decision trees predict 1, and 2 predict 0. Hence, prediction 1 will be considered.



Swipe!

RANDOM FOREST

- **The underlying principle of a random forest is that several weak learners combine to form a keen learner.**

- **The steps to build a random forest are as follows:**

- Build several decision trees on the samples of data and record their predictions.
- Each time a split is considered for a tree, choose a random sample of mm predictors as the split candidates out of all the pp predictors. This happens to every tree in the random forest.
- Apply the rule of thumb i.e. **split m = p$\sqrt{m}$ =p**
- Apply the predictions to the majority rule.

@insaid.learn

Swipe!

INSAID

WHAT IS DEEP LEARNING?

@insaid.learn

DEEP LEARNING

- It is a paradigm of Machine Learning. In deep learning, multiple layers of processing are involved in order to extract high features from the data.

- The neural networks are designed in such a way that they try to simulate the human brain

- Deep Learning has shown incredible performance in recent years because of the fact that it shows great analogy with the human brain.

@insaid.learn

Swipe!

PREVIOUS POST:

@insaid.learn

Swipe!

SHARE IT & SAVE IT

FOR FUTURE REFERENCE

@insaid.learn

FOLLOW FOR MORE

Follow us @insaid.learn