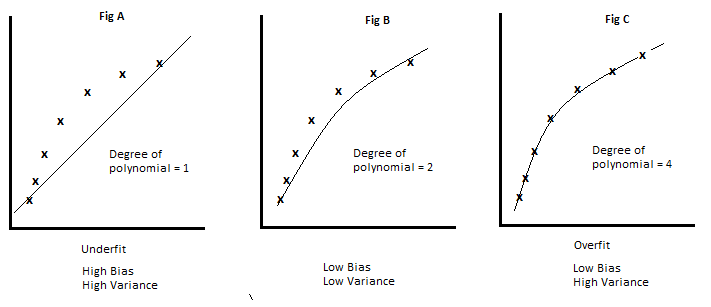
Bias and Variance



Bias = Error in training data

Variance = Error in testing data

You’re trying to create a best fit line using a model, and as you see in:

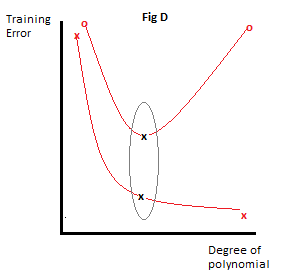
Fig A, where there is under fitting, the model has high bias (high error in training data) therefore it will also have high variance(high error in testing data)

Fig C, when there is overfitting, the model will have low bias (low error in training data) but will have high variance (high error in testing data)

Fig B, is a generalized model, with low bias and low variance.

Therefore, if you have models as below:

|  |  |  |
| --- | --- | --- |
| Model 1 | Model 2 | Model 3 |
| Training Error = 1% | Training Error = 25% | Training Error < 10% |
| Test error = 20% | Test error = 26% | Test error < 10% |
| Overfit model | Underfit model | Generalized model |



If we plot a graph of training error vs degree of polynomial,

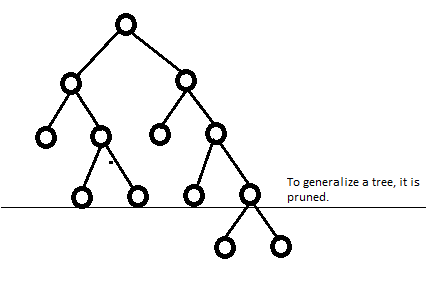
The x,o on the left are for fig A, it has a high error and degree of polynomial 1, similarly x,o on the right is for Fig C, with degree of polynomial 4, and high variance.

The variance drops to a certain extent, as the degree of polynomial increases, beyond which it starts increasing again.

We have to generalize the model to the point x,x where the degree of polynomial and error is optimal.

# Decision Tree vs Random Forest

1. Decision Tree



A decision tree if let grow without constraint is an example of overfit. It has low Bias but high variance. Therefore we always have to prune decision trees.

Decision trees is a basis for Bagging and Boosting.

Random forest is an example of Bagging, and XG boost is an example of bootstrapping.

Random Forest: Performs **Bootstrap Aggregation**

So, in order to resolve the high variance issue with decision trees, we use Random Forest.

In Bootstrap aggregation, data is split and given to different decisions trees, each decision tree has a capability of low bias and high variance (Overfitting), but when you combine them in parallel and aggregate the results, the high variance gets converted to low variance. Therefore you get a generalized model with low bias and low variance.

