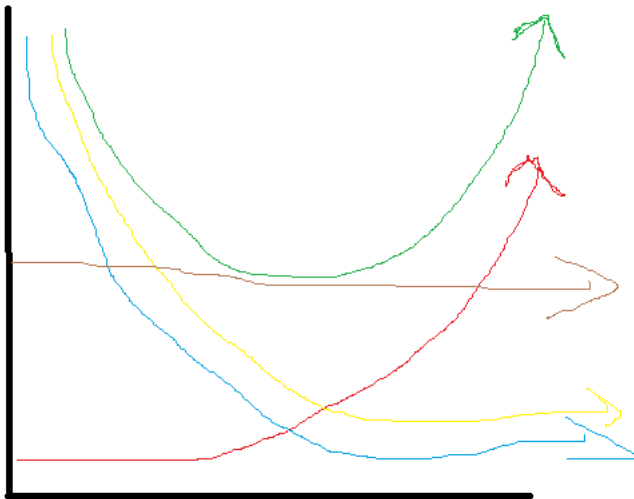


Assignment4

John Bute

2024-09-25

Assignment 4



Legend:

- Brown: irreducible error
- Yellow: training error
- Green: test error
- Red: Variance
- Blue: Bias
- X axis: amount of flexibility in the method
- Y axis: values of each curve

Bias: Bias is the difference between the expected predictions and the true values. Thus, simpler statistical learning methods end up underfitting the model, which cause predictions to be far off actual values of the dataset, displaying high bias. Meanwhile, a complex statistical learning method that fits the data better will hit the mark more often.

Variance: Variance refers to the statistical learning method's sensitivity to fluctuations of data points within a dataset. When we retrain models on different subsets of the data, the changes in prediction of the statistical learning method is variance. A simpler method has low variance, as they do not capture fluctuations or noise in data as much as complex data do. As a result, the method is barely affected when there is a change in the data. However, the more you increase the statistical learning method's complexity, the more the method learns the fluctuations in training data. However, if the method becomes too complex, the model may overfit, causing high variance, as when you alter the dataset, the model will greatly change to accommodate the altered data, thus changing the method's prediction greatly.

Training Error: When you first train a statistical learning method, you train on a training set. Simpler method tend to have a high training error, as you are underfitting the data, meaning you are not representing a general pattern in the data well, resulting in a high error rate. The more complex your method becomes, the more it fits your training set, recognizing the pattern between inputs and outputs. Eventually, the training error will be very low, as a highly complex model will overfit the dataset, i.e. the method fits exactly to its training data, causing the method to make highly accurate predictions on the training set.

Test error: If you train a highly complex method or a very simple method, you will end up with a high test error, as your method is not generalizing the data correctly. You need to strike a balance between underfitting and overfitting your method, thus allowing a minimal test error. Minimizing test error often requires finding the right balance between variance and bias.

Bayes (irreducible) error: Irreducible error consists of errors caused by noise or randomness. It is the lowest error achievable, thus it is a rate that we cannot reduce from the complexity of the model as it is entirely based on the dataset. As a result, it remains constant from simpler method to complex method.