Overfitting: Takeaways ⤴

Concepts

- Bias and variance are at the heart of understanding overfitting.
- Bias describes error that results in bad assumptions about the learning algorithm. Variance describes error that occurs because of the variability of a model's predicted values.
- We can approximate the bias of a model by training a few different models using different features on the same class and calculating their error scores.
- To detect overfitting, you can compare the in-sample error and the out-of-sample error, or the training error with the test error.
    - To calculate the out-of-sample error, you need to test the data on a test set of data. If you don't have a separate test data set, you can use cross-validation.
    - To calculate the in-sample-error, you can test the model over the same data it was trained on.
- When the out-of-sample error is much higher than the in-sample error, this is a clear indicator the trained model doesn't generalize well outside the training set.

Resources

- Bias-variance tradeoff
- Blog post on the bias-variance tradeoff