Regularization and the Bias-Variance Tradeoff

TM Quest

Overview

What Will we Learn in This Module?

- What is regularization (or shrinkage)?
  - How to implement Ridge regression.
  - How to implement Lasso regression.
  - What are the advantages of regularization?
- What is bias and variance?
  - What is the bias-variance tradeoff?
  - How should the bias-variance tradeoff affect our decisions?

Regularization (or Shrinkage)

Motivation

Linear Regression Advantages

- It is simple to understand.
- It runs quickly.
- It is easy to interpret the results.

Linear Regression Disadvantage

- It often performs worse than a complex model such as a random forest.

Solution: Improve linear regression!

Regularizing Linear Regression

The Idea of Regularization

The idea behind regularization is to shrink the parameters in a linear regression model so that the most prominent parameters stand out.

Say you have a linear regression model

$$a_1x_1 + a_2x_2 + a_3x_3 = y,$$

and when fitting it to your dataset you get

$$1.7x_1 + 0.2x_2 + 3.1x_3 = y.$$

Setting $a_2 = 0$ gives the regularized model

$$1.7x_1 + 3.1x_3 = y.$$

More Sophisticated Regularization

Lasso and Ridge Regression

Lasso regression and Ridge regression are two regularization techniques for linear regression.

Facts

- Similarities
  - Are built into scikit-learn and are easy to use.
  - Have a hyperparameter that determines the level of shrinkage.
- Differences
  - Lasso regression will typically set some of the parameters to zero.
  - Ridge regression shrinks parameters but doesn't set them to zero.

Bias and Variance Tradeoff

Explaining Bias and Variance

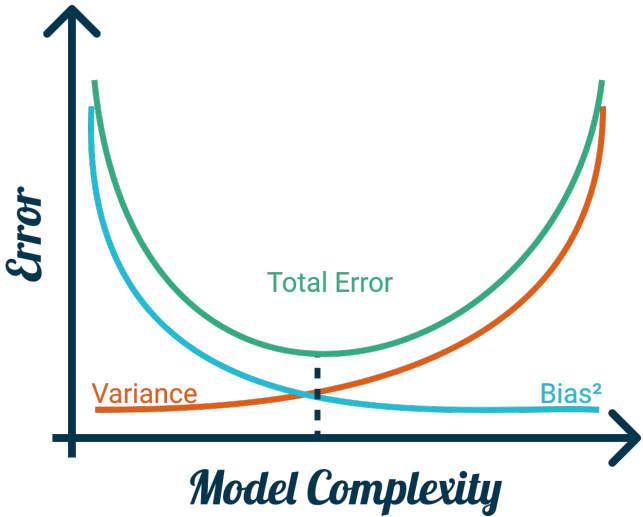
Bias

- The bias error is an error arising from wrong assumptions in the model.
- An example of high bias error is using a linear model when the data is highly non-linear (underfitting)

Variance

- The variance error is an error arising from sensitivity to small fluctuations in the training set.
- An example of high variance error is using a complex model that picks up random noise in the training set (overfitting).

Bias-Variance Tradeoff



- Our goal is to reduce the total error.
- Want to simultaneously reduce the bias error and variance error as much as possible.
- The bias-variance tradeoff states that lowering the bias error often increases the variance error and vice versa.

Bias-Variance in Regularized Models

The α Hyperparameter

Lasso regression and Ridge regression have a hyperparameter α.

- When $\alpha = 0$ we have the usual linear regression.
- When $\alpha \to \infty$ all the parameters go to zero.

What is happening?

- Increasing α lowers the variance error but increases the bias error.
- Decreasing α increases the variance error but decreases the bias error.
- The best α finds a sweet spot where the total error is lowest.