Statistical Learning - Homework 4 (Conceptual)

James Hahn

Chapter 5 - Exercise 4

We can use the bootstrap to estimate the standard deviation of our predictions. Let's say we're given some sample (X) of size N from the population (P). Normally, in order to estimate the std, we could just sample the population a large number of times and view the distribution of the std after sampling over and over again. However, this is infeasible since we only have access to a certain subset of the data, X. As such, we will take a random sample (X') of X of size N (same size as the sample) with replacement (this is called a bootstrap sample), calculate its standard deviation, and repeat this process B times, each time generating a new bootstrap sample and calculating a new standard deviation on X'. Those B standard deviations are represented as $\hat{\alpha}_1 \dots \hat{\alpha}_b$. The process we will be using is bagging (bootstrap aggregating). In order to calculate the estimated standard deviation, we can take the average of the std. estimates as follows:

$$\bar{\alpha} = \frac{1}{B} \sum_{i=1}^{B} \hat{\alpha}_i$$