Lesson 6.02
Bootstrapping & Bagging

Bootstrapping

- Bootstrapping resamples the original dataset with replacement several times to create simulated datasets.

- This process involves drawing random samples of the same size from the original dataset

Bootstrapping Process

1. The bootstrap method has an equal probability of randomly drawing each original data point for inclusion in the resampled datasets.

2. The procedure can select a data point more than once for a resampled dataset. This property is the "with replacement" aspect of the process.

3. The procedure creates resampled datasets that are the same size as the original dataset.

Bootstrapping

- The process ends with your simulated datasets having many different combinations of the values that exist in the original dataset.

- Each simulated dataset has its own set of sample statistics, such as the mean, median, and standard deviation.

- Bootstrapping procedures use the distribution of the sample statistics across the simulated samples as the sampling distribution.