DATA 100: Vitamin 11 Solutions

April 26, 2019

1 Random Forests

1.1 Definition

To train a Random Forest regression model, fit a ___ to each of several ___ and select covariates at random for consideration at each node.

- ☐ linear regression model, randomly sampled learning sets
- ☐ regression tree, randomly sampled learning sets
- ☐ linear regression model, bootstrap samples of the learning set
- ☑ regression tree, bootstrap samples of the learning set

1.2 By-Products

Which of the following are useful by-products of Random Forests?

- ☑ variable importance measures
- ☑ risk estimates
- ☑ observation proximity measures
- ☐ an estimate of model bias

2 Statistical Inference

2.1 Sampling Distribution

A sampling distribution is:

- ☐ A function of a sample that estimates the true value of a parameter
- ☐ The observed empirical distribution of an estimator's values based on a sample

- ☑ The unknown distribution of an estimator's value for different random samples from the population
- ☐ The observed distribution of an estimator's value for different bootstrap samples

2.2 Bootstrap Assumption

Select all that apply: Bootstrap resampling can be used to:

- ☑ Estimate parameters of a sampling distribution
- ☑ Estimate the variance of an estimator
- ☑ Estimate a confidence interval for an estimator
- ☑ Generate samples that resemble random samples from the population, assuming that the bootstrap population is representative of the population.

2.3 Confidence Interval Interpretation

Suppose we compute the 95% bootstrap confidence interval of some unknown parameter θ to be (a, b) using a random sample of size n. Which of the following should be true?

- ☐ There is a 95% chance that the true value of θ falls within (a, b).
- ☐ The probability that the true value of θ is within the 95% confidence interval is 95%.
- ☑ About 95% of the estimates computed from bootstrap samples fell within (a, b).
- ☑ Under multiple repetitions of the bootstrap procedure, we would expect the true value of θ to be within the 95% confidence intervals 95% of the time.