

Statistical/Machine Learning Methods

Zach Ginder

Task 1: Conceptual Questions

- Question 1: What is the purpose of using cross-validation when fitting a random forest model?
- Question 2: Describe the bagged tree algorithm.
- Question 3: What is meant by a general linear model?
- Question 4: When fitting a multiple linear regression model, what does adding an interaction term do? That is, what does it allow the model to do differently as compared to when it is not included in the model?
- Question 5: Why do we split our data into a training and test set?

Task 2: Data Prep