Day 3 Morning: Machine Learning

Gabriella Gerlach

December 10, 2020

1   Data munging for ML

Associated screen cast: link Load the fmri data using `seaborn.load_datasets("fmri")`

1. Remove the column "subject"

2. Convert the categorical variables to dummies

3. Extract "signal as the outcome of interest

2   Data splitting

Associated screen cast: link Split the dataset from the previous assignment into a 80/20 training/test split

3   fit the model

Associated screen cast: link Fit a decision tree and a random forest model to the training data

4   model performance

Associated screen cast: link Determine the performance of the models using mean squared errors and R2

5   Cross-validation

Associated screen cast: link Perform 5-fold cross-validation (split into 5 pieces) to better estimate the errors. compare to prediction error using test data.

Draw a graph to show how well each model predicted an individual's signal