Lecture 27 Outline:

1. Random Forests.
   (a) Review Random Forest algorithm.
   (b) Comparisons to Boosting & Bagging.
   (c) Bias unchanged.
   (d) Variance decreases more than that of bagging.
   (e) How many split variables?
   (f) Can idea be used for other learning models?
   (g) OOB error estimates & number of bootstrap samples.
   (h) Variable Importance.
   (i) Overfitting?
   (j) Relationship to $K$-nearest neighbors.

2. Ensemble Building.
   (a) Combining Random Forest ideas with forward stage-wise modeling.
   (b) Model Stacking.
      i. Learning optimal weights.
      ii. Post-processing weights via shrinkage.
   (c) Consensus voting reduces variance.
   (d) Less correlated learners perform best in ensembles.

Reading Assignment:

• Textbook, Chapter 15 & 16.