Lecture 22 Outline:

1. Classification and Regression Trees (CART):
   (a) Binary recursive partitioning.
   (b) Tree structures, nodes, splits.

2. Regression Trees:
   (a) Squared error criterion.
   (b) Computing the splits.
   (c) Recursive algorithm.

3. Classification Trees:
   (a) Partition based on proportions in each class.
   (b) Node impurity measures:
      - Missclassification Error.
      - Gini Index. (Bernoulli variance).
      - Cross-Entropy.

4. Size of Trees.
   (a) Over- and under-fitting: similarities to other methods.
   (b) Cost-Complexity pruning.
   (c) Unique sequence of nested trees.
   (d) Criterion to determine the optimal tree size.

5. Categorical variables.
   (a) Order according to mean response or proportion in each class.

   (a) Treat missing as a separate category.
   (b) Surrogate splits.

7. Real data example.

8. Strengths & Weaknesses of Trees.