Chapter 3 Correlation to Supervised Segmentation

Predictive model v.s. descriptive model

Descriptive:
- gain insight into the underlying phenomenon or process
- E.g. what do customers who churn look like?

Predictive:
- Whether a certain customer will churn or not

Supervised learning models:
- Features/attributes
- Target
    o Values of target: label
- Dataset/tables
- Instances/example/observation
- Training data / labeled data: because label is known

Selecting Informative attributes

Intuition: partition dataset into subsets based on different values of one attribute, to see whether target in the group is "pure" or "mixed"

Entropy
- measure of disorder in a group

Information Gain (IG)
- Parent set's entropy - weighted overall entropy of a segmentation of an attribute
- Weighted refers to entropy of a specific group (based on a value of an attribute), times proportion of population of the group, and sum all groups together in this way

Attributes with high IG will be more useful to us to predict target

Presenting IG by graph:
- X axis: proportion of population in each partition
- Y axis: entropy of each partition, the ceiling is entropy of total population
- Shaded area: weighted entropy of that attribute. The blank is reduce of entropy

Tree structured models
- Want to select multiple attributes and combine them
- On each decision node, select attribute with most IG (conditional on parent node) to be the next one of division
- Classification tree can be used for binary variables, probability estimation (output as portion of subgroup, Laplace correction if necessary), or regression tree for numeric. The idea is the same