Nonmetric Methods

Chapter 8 Machine Learning

Decision Trees

- When the parameters that constitute a sample aren't merely numerical/ metrics, we can use decision trees to classify.
- With that being said, you can still use decision trees on metric data.
- This introduces the concept of impurity because the boundaries generated can produce contaminant samples within each classifier.
    - An impurity of 0 for a node means that it's 'pure' and contains only 1 class
    - Large impurity suggests a lot of mixing of classes between nodes.
    - A common metric is the Gini impurity for a node N

$$i(N) = \sum_{i \neq j} P(w_i)P(w_j) = 1 - \sum_j P^2(w_j)$$

    - Essentially we are looking at the probability of each class being present at a node - the more classes present, the more we pollute this impurity. In the extreme pure case that P(w_1) = 1 and everything else is 0, you get i(N) = 1-1 = 0. Other cases, your second term is <1 so you get i >0
    - When optimising based on impurity (e.g. via gradient descent) it is important not to necessarily get an impurity of 0 as this suggests overfit

Random Forests

- Uses the majority vote of many trees to come to a consensus on classification