Lecture 11 Decision Trees

* Decision trees can be used to predict classes based upon a yes no pathway
* A diagram of a relationship

  Description automatically generated
* Hunt’s Algorithm
  + Hunt’s Algorithm is a classic recursive decision tree learning algorithm. It builds a decision tree by repeatedly splitting the dataset based on attribute values to create pure subsets (i.e., subsets where all instances belong to the same class).
  + Base Cases:
    - If split and all data points in the same class
      * Predict that class
    - If split and no data points
      * Predict a reasonable class
  + Splitting
    - Choosing the attribute and condition that most effectively divies the dataset into purer subsets – groups where the data points mostly belong to the same class
    - Goal: reduce impurity at each data point. A good split results in the same classes represented amoungst data points
    - Binary Split
      * Split the attribute into two groups
      * Ex age > 30 and age< 30 or weather = sunny vs, weather != sunny
    - Multi-way split
      * The attribute is split into multiple groups one group for each unique value
      * Ex. Attribute = weather w categories sunny, rainy or overcast -> 3 branches for the split
  + GINI Index
    - Metric used to evaluate how “pure” a dataset is after a split in a decision tree algorithm
    - A math equations and numbers

      Description automatically generated with medium confidence
  + GINI of the Split
    - nt= number of data points at node t
    - n = number of data points before the split (parent node)
    - A diagram of a number of numbers

      Description automatically generated with medium confidence
  + Limitations
    - Easy to construct a tree that is too complex and overfits the data
    - Solutions
      * Early termination : stop before the tree is fully grown – use a majorirt vote at the leaf node
        + Stop at some specific depth
        + Stop if size of node is below some threshold
        + Store if gini does not improve
      * Pruning : trim tree based on assigned values
  + Extensions

A math equations on a white background

Description automatically generated