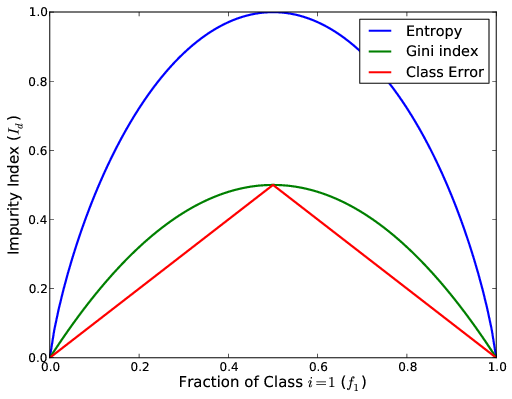
**Text**

**Data**

**Text**

**Decision Trees Ch4**

**# Impurity measures(I): always I(pure) = 0**

****

**Entropy:**

Fair coin:

**Entropy(0.5) =**

Biased coin to give 99% head:

**Entropy(0.99) =**

**Gini:**

Fair coin:

**Gini(0.5) =**

Biased coin to give 99% head:

**Gini(0.99) =**

**Classification error:**

Fair coin:

**CE(0.5) =**

Biased coin to give 99% head:

**CE(0.99) =**

**Remainder=**

**# Information Gain =** if root I(parent) =1

**I** is the impurity measure of a given node, **N** is the total number of records at the parent node, **k** is the number of attribute values, and **N(vi)** is the number of records associated with the child node **vi**.

**Example** binary classification with 12 records in the training dataset, we choose I = entropy

|  |  |  |  |
| --- | --- | --- | --- |
| record | color | Make | class |
| 1 | Red | GMC |  |
| 2 | Red | Jeep |  |
| 3 | Green | GMC |  |
| 4 | Green | Jeep |  |
| 5 | Green | Toyota |  |
| 6 | Green | Ford |  |
| 7 | Blue | Ford |  |
| 8 | Blue | Ford |  |
| 9 | Blue | Toyota |  |
| 10 | Blue | Toyota |  |
| 11 | Blue | Ford |  |
| 12 | Blue | Toyota |  |

Gain(Color) =

Gain(Make) =

**So, we split on color**

**# Split information =**

if each attribute value has the same number of records (equal partitions) then and split information would be **log k**

**# Gain ratio:** Impurity measures tend to favor attributes that have a large number of distinct values. A test condition that results in a large number of outcomes may not be desirable because the number of records associated with each partition is too small to enable us to make any reliable prediction.

Decision Tree Algorithm

1. If the remaining examples of a node are pure, then done.
2. If there are no attributes left, but only classifications, means that these examples have exactly the same description but different classification. This can happen because there is an error or noise in the data. Return the majority voting. (plurality classification of the remaining examples).
3. If there are no examples left, it means that no examples has been observed for this combination of attribute values. Return the vote of the parent. we return a default value calculated from the plurality classification of all examples that were used in constructing the parent.
4. Else choose the best attribute to split. Entropy for each value, then sum proportionally (remainder), then information gain (1 – remainder)

**Ensemble methods Ch5**

**Text**

**Bayesian Classifier Ch5**

**Text**



**=**

**= =**

**Association Analysis Ch6 Ch7**

**Clustering Concepts Ch8**

**Text**

**More**

**Text**