# Zaki - CH19 - Decision Trees

$D$ = trainig set of $n$ points of $x$ in a $d$-dimensional space

$y_i$ = ~~the~~ class label of $x$

$X_j$ = Attributes of $X$

$K$ = number of Distinct classes

$y_i = \{c_1, c_2, \ldots, c_n\}$

A decsion tree classifier is a recursive, partition based tree model that predicts the class $\hat{y}_i$ for each point $x_i$

$R$ = data space that the trainig set $D$ is a part of

A decsion tree uses an axis parrell hyperplane to split the data space $R$ to half spaces/regions, $R_1$ & $R_2$

this also partitions the train data into $D_1$ & $D_2$

Each of these regions is recursively split via axis-parralleI hyperplanes until the points X's with each region is relatively pure in terms of their class labels

the resulting hierachy of split decsions creates a decision tree model

The leaf nodes are labelled w/ the majority class among points the the region

New points can be passed through the model tree until it reaches a leaf Node

the prediction = the Class of the leaf node

# Axis-Parallel Hyperplanes

A hyperplane $h(x)$ is defined as the set of all points $x$ that satisfy:

$$h(x) = w^T x + b = 0$$

$w \in \mathbb{R}^d$ is a weight vector that is normal to the hyperplane

$b$ is the offset of the hyperplane from the origin

The weight vector must be parrel to one of the orginal axes $X_j$

↓

This means that one of the values will be 1 & the rest 0

This creates a sort of selector functionality where we choose the attribute to split by

An example of a ~~a~~ hyperplane:

$h_1(x): x_1 - 5.45 = 0$

Determing and organsias
Split

$\leq 0$ or $> 0$

---

## Parity

Parity is the fraction of points w/ the majoriy label $D_j$

$$\text{Parity}(D_j) = \max_i \left\{ \frac{n_{ji}}{n_j} \right\}$$

# Split Point Evaluation Measures

Split forms:
num: $X_j \le v$ cat: $X_j \in V$

We need an objective criterean for scoring the split point

want to select the point point that gives the best separation ~~of~~ ~~discrimination~~ within classes

the two main types are
Entrophy & Gini index
↓
info Gain

# Entrophy

In general, Entrophy measures the amount of disorder or un-certainty in a system

In classification, if entropy is low then the split is more pure

higher = more disorder = Class labels are mixed & no majority class

Entrophy of D labeled points:

$$H(D) = -\sum_{i=1}^{k} P(c_i \mid D) \log_2 P(c_i \mid D)$$

$P(c_i \mid D)$ probability of class $c_i$ in $D$
$k$ = num of classes

Assume a split parts into $D_Y$ & $D_N$

Define as weighted entrophy of the resulting partitions:

$$H(D_Y, D_N) = \frac{n_Y}{n} H(D_Y) + \frac{n_N}{n} H(D_N)$$

To see if split results in reduced entrophy, we define info gain for the split as:

$$Gain(D, D_Y, D_N) = H(D) - H(D_Y, D_N)$$

higher the gain = more reduced entrophy = better the split

# Gini Index

$$G(D) = 1 - \sum^{K} P(c_i|D)^2$$

if Spit is pure then $= 1$

$1^2 = 1 \quad , \quad 1 - 1 = 0$

$0^2 = 0 \quad , \quad 1 - 0 = 0$

All other class will be $\cancel{0^2 = 0, \ 1 - 0 = 0}$

hence the gini $= 0$, $1 - \sum(1^2, 0^2) = 0$

If all values are equally weight the Prob $P(c_i|D) = 1/k$ then the Gini $=$

$$\frac{k-1}{k}$$

Thus, higher gini = More disorder
lower gini is better

Weighted Gini

$$G(D_Y, D_N) = \frac{N^Y}{N} G(D_Y) + \frac{n_N}{N} G(D_N)$$