Decision Trees

Decision Tree

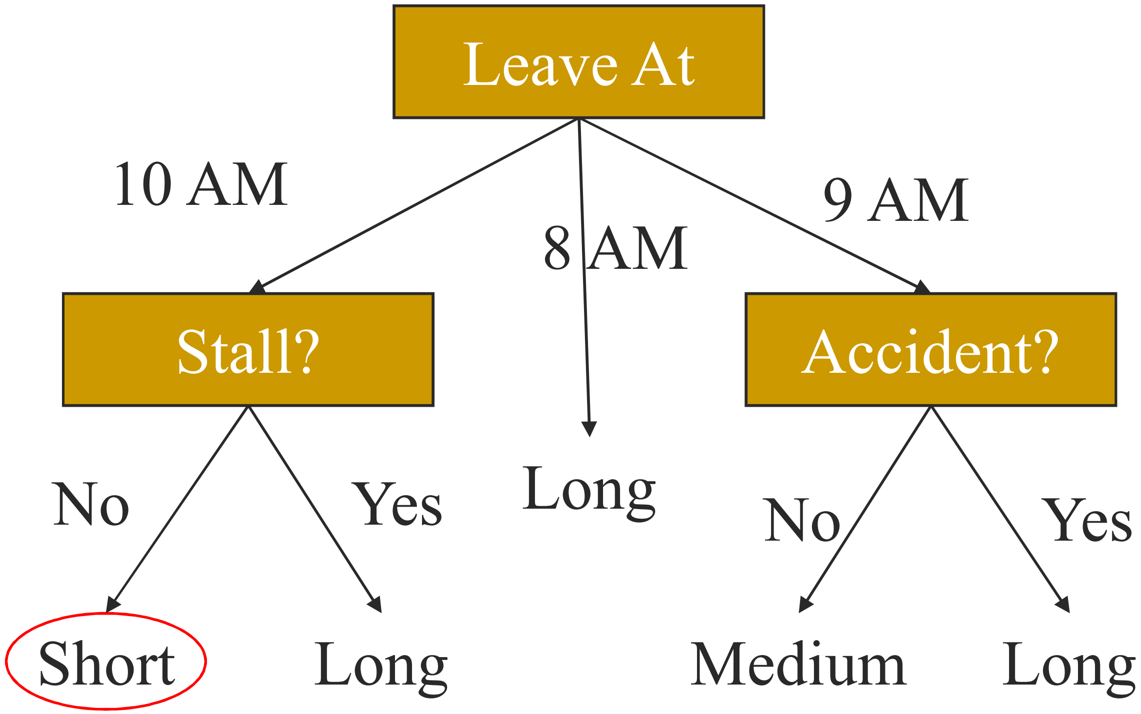
- Python:

  https://www.anaconda.com/download/

- Pycharm:

  https://www.jetbrains.com/student/

Decision Tree



If we leave at 10 AM and there are no cars stalled on the road, what will our commute time be?

Decision Tree as a Rule Set

if hour == 8am

    commute time = long

else if hour == 9am

    if accident == yes

        commute time = long

    else

        commute time = medium

else if hour == 10am

    if stall == yes

        commute time = long

    else

        commute time = short

- Notice that all attributes to not have to be used in each path of the decision.
- As we will see, all attributes may not even appear in the tree.

ID3

- Calculation of entropy
  - $\text{Entropy}(S) = \sum_{(i=1 \text{ to } l)} -|S_i|/|S| * \log_2(|S_i|/|S|)$
    - S = set of examples
    - S_i = subset of S with value v_i under the target attribute
    - l = size of the range of the target attribute

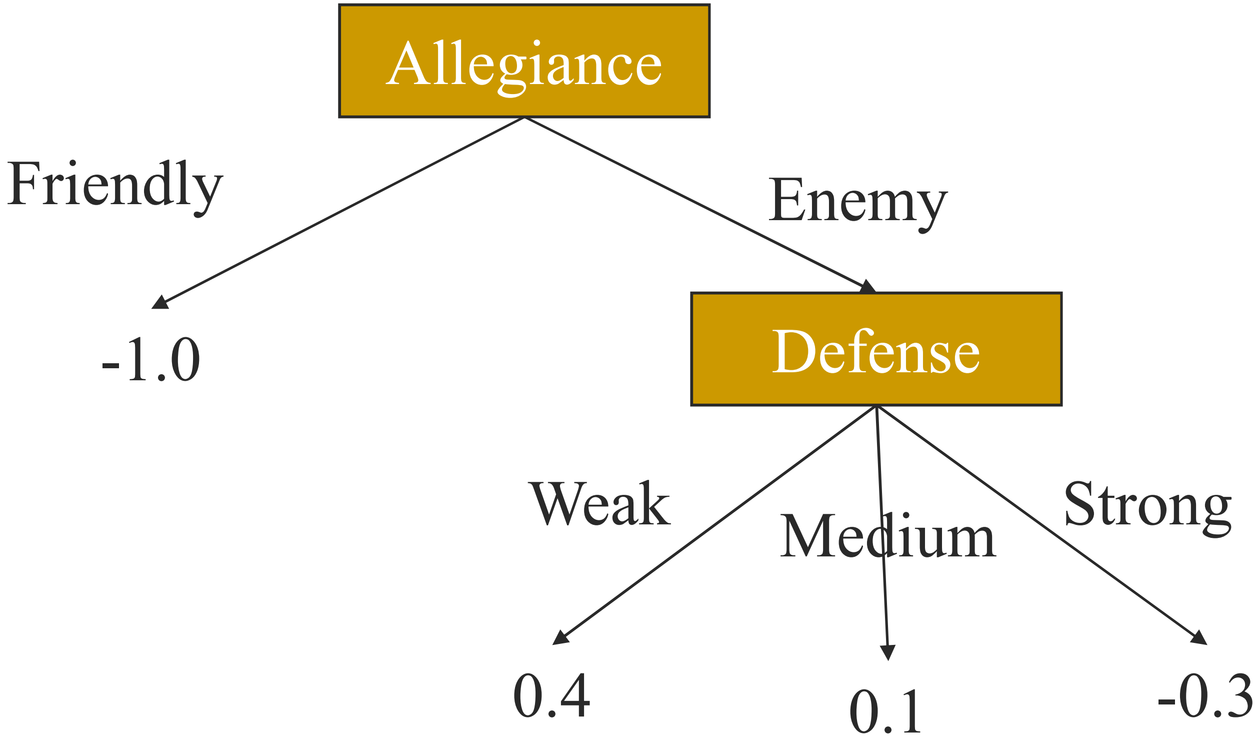
ID3

- ID3 splits on attributes with the lowest entropy
- We calculate the entropy for all values of an attribute as the weighted sum of subset entropies as follows:
  - $\sum_{(i = 1 \text{ to } k)} |S_i|/|S| \text{ Entropy}(S_i)$, where k is the range of the attribute we are testing
- We can also measure information gain (which is inversely proportional to entropy) as follows:
  - $\text{Entropy}(S) - \sum_{(i = 1 \text{ to } k)} |S_i|/|S| \text{ Entropy}(S_i)$

ID3 in Black & White

Example	**Attributes**			**Target**
	Allegiance	Defense	Tribe	Feedback
D1	Friendly	Weak	Celtic	-1.0
D2	Enemy	Weak	Celtic	0.4
D3	Friendly	Strong	Norse	-1.0
D4	Enemy	Strong	Norse	-0.2
D5	Friendly	Weak	Greek	-1.0
D6	Enemy	Medium	Greek	0.2
D7	Enemy	Strong	Greek	-0.4
D8	Enemy	Medium	Aztec	0.0
D9	Friendly	Weak	Aztec	-1.0

ID3 in Black & White



C4.5

- What’s the difference?

$$g_R(D,A) = \frac{g(D,A)}{H(D)}$$

CART

- Splitting point

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right]$$

$$\begin{cases}\hat{c}_1=\text{ave}(y_i\mid x_i\in R_1(j,s))\\ \hat{c}_2=\text{ave}(y_i\mid x_i\in R_2(j,s))\end{cases}$$

Pruning

- Loss function

$$C_\alpha(T) = C(T) + \alpha |T|$$

$$g(t) = \frac{C(t) - C(T_t)}{|T_t| - 1}$$

$$\alpha = \min(\alpha, g(t))$$