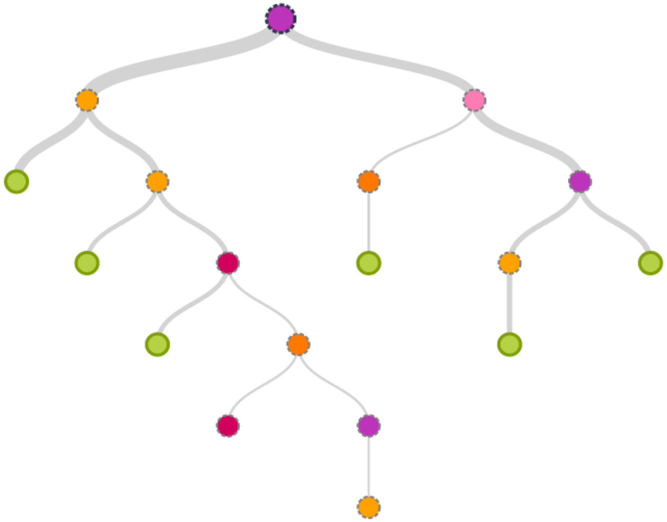
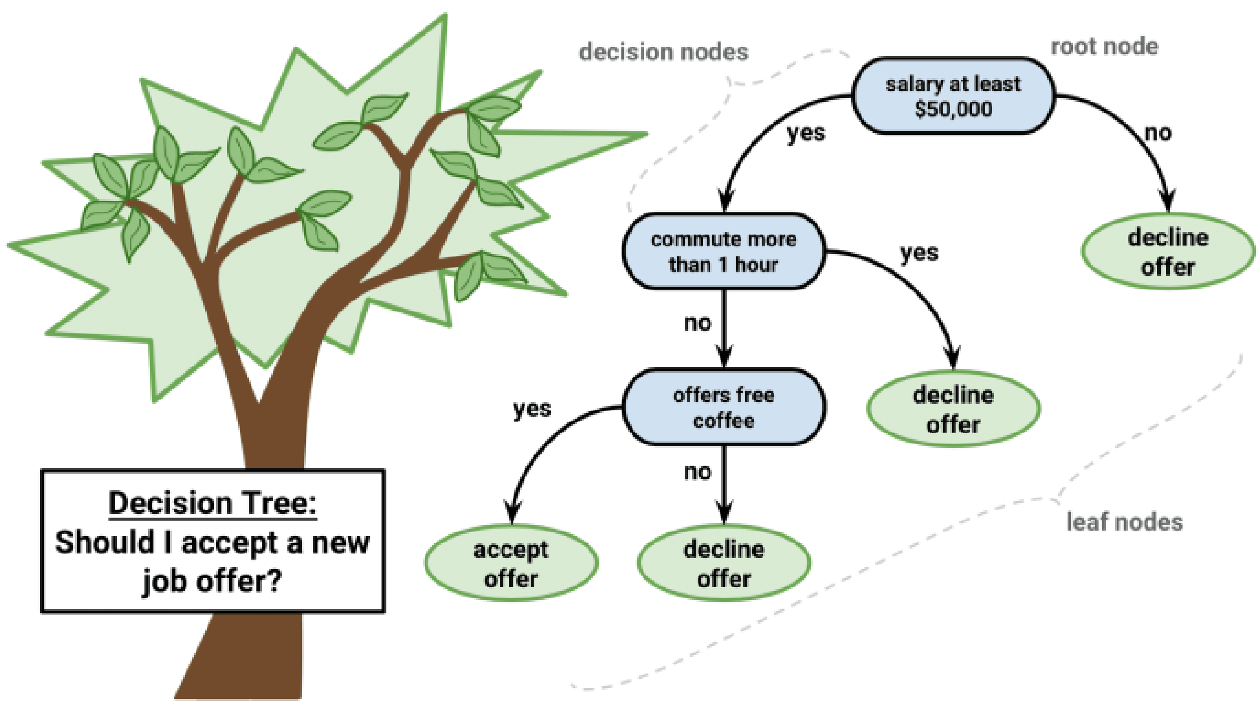
Decision Tree



Nabila Shahnaz Khan

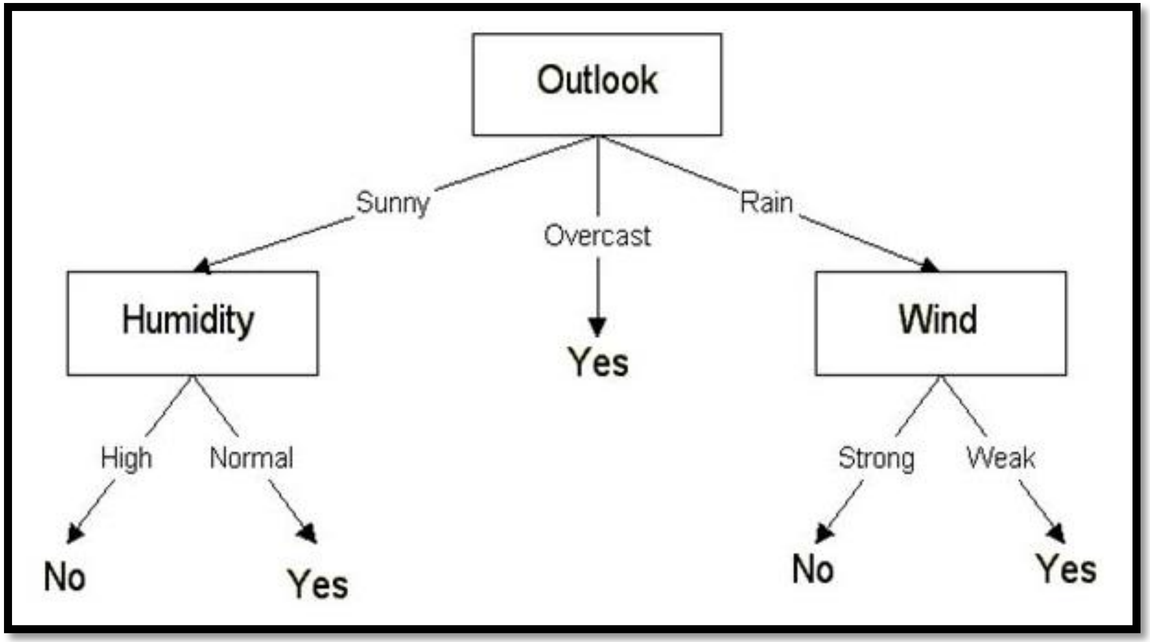
What is Decision Tree?

A decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision.



Decision Tree

- A decision tree is drawn upside down
- It's internal nodes represent the features of the classifier
- Each link branch represents a decision (rule)
- The external or leaf nodes represent the decision



Decision Tree

Decision trees are used for both **classification** and regression problems

Why Decision Trees?

- Decision tress often mimic the human level thinking so its so simple to understand the data and make some good interpretations.
- Can Handle both continuous and discrete data
- Decision trees are "white boxes" in the sense that the acquired knowledge can be expressed in a readable form, while KNN,SVM,NN are generally black boxes, the acquired knowledge can not be read in a comprehensible way. So can be easily visualized and manipulated.

Sample Dataset

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes

outlook	temp.	humidity	windy	play
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Question is

Which feature to select first, 2^{nd} and so on???

Different Decision Tree Algorithms

- **CART** (Classification and Regression Trees) → uses Gini Index Classification as metric.
- **ID3** (Iterative Dichotomiser 3) → uses Entropy function and Information gain as metrics.
- C4.5 (Successor of ID3) → Similar to ID3, can handle both discrete and continuous values (using threshhold) and missing attribute values.
- Others: CHAID (CHi-squared Automatic Interaction Detector), MARS (Multivariate Adaptive Regression Splines), Conditional Inference Trees

ID3 DT Algorithm

Steps:

1. Compute the entropy for dataset

S=Dataset, C={yes, no}

Entropy of Weather Dataset, H(S)=$\sum_{c \in C} -p(c)\log_2 p(c)$

Ex:

p(yes)= -(9/14)*$\log_2\left(\frac{9}{14}\right) = 0.41$

p(no)= -(5/14)*$\log_2(\frac{5}{14}) = 0.5$

H(S)=p(yes)+p(no) = 0.94

ID3 DT Algorithm

Steps:

2. For each feature:

a) calculate entropy(E) for all categorical values of it and take their average; this is the information entropy(E) of current attribute

b) calculate information gain(I) for the current attribute

E(Outlook=sunny)= $-\frac{2}{5} * \log_2\left(\frac{\mathbf{2}}{\mathbf{5}}\right) - \frac{3}{5} * \log_2\left(\frac{\mathbf{3}}{\mathbf{5}}\right) = 0.971$

E(Outlook=overcast)= $-1 * \log_2(\mathbf{1}) - 0 * \log_2(\mathbf{0}) = 0.0$

E(Outlook=rainy)= $-\frac{3}{5} * \log_2\left(\frac{\mathbf{3}}{\mathbf{5}}\right) - \frac{2}{5} * \log_2\left(\frac{\mathbf{2}}{\mathbf{5}}\right) = 0.971$

Average Entropy of Outlook, E(Outlook)= $\frac{5}{14} * 0.971 + \frac{4}{14} * 0 + \frac{5}{14} * 0.971$

Information Gain, I(Outlook) = E(S) – E(Outlook) = 0.94–0.693=0.247

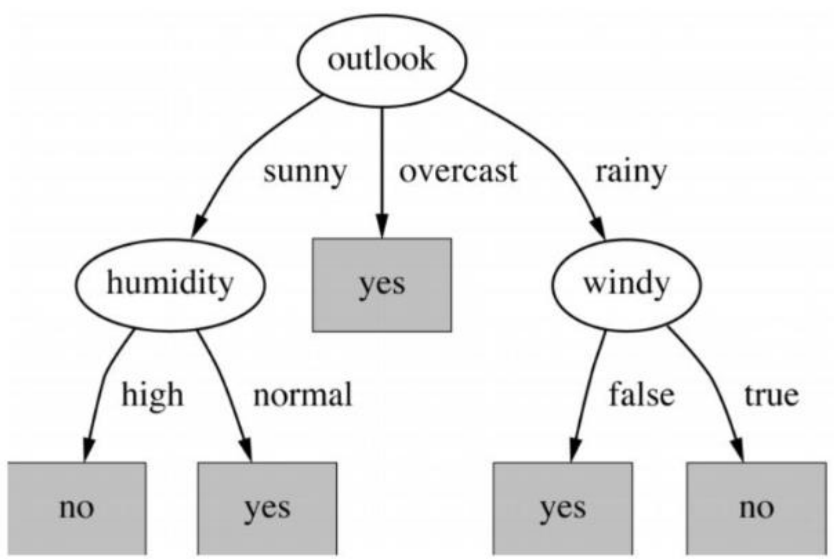
*****Similarly Calculate Information gain for all other features**

ID3 DT Algorithm

Steps:

3. Pick the feature with the highest information gain and take it as root of the decision tree.

4. Repeat this procedure for all other remaining features and select the other features as internal nodes one by one. Repeat until the complete decision tree is formed.



CART DT Algorithm

Steps:

1. Compute the gini index for data-set.

S=Dataset, C={yes,no}

Gini index of Weather Dataset, Gini(S)= 1- $\sum_{c \in C} {P_c}^2$

$$\text{Gini(S)} = 1 - \left(\frac{9}{14}\right)^2 - \left(\frac{5}{14}\right)^2$$
$$= 1 - 0.413 - 0.127$$
$$= 0.46$$

CART DT Algorithm

Steps:

2. For each feature:

a) calculate gini index for all categorical values of it and take their average

b) calculate gini gain(I) for the current attribute

Gini(Outlook=sunny)=1 - $(\frac{2}{5})^2$ - $(\frac{3}{5})^2$ = 0.48

Gini(Outlook=overcast)= 1 - $(\frac{4}{4})^2$ - $(\frac{0}{4})^2$ = 0

Gini(Outlook=rainy)= 1 - $(\frac{3}{5})^2$ - $(\frac{2}{5})^2$ = 0.48

Average Gini index of Outlook, Gini(Outlook)=$\frac{5}{14} * 0.48 + \frac{4}{14} * 0 + \frac{5}{14} * 0.48$

Gini Gain, I(Outlook) = 0.46 – Gini(Outlook) = 0.46–0.34=0.12

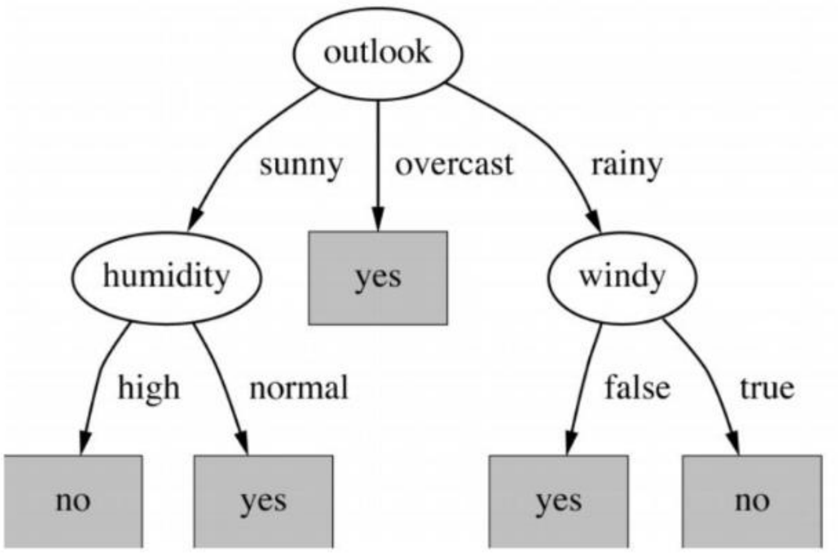
*****Similarly Calculate Gini gain for all other features**

CART DT Algorithm

Steps:

3. Pick the feature with the best gini gain and take it as root of the decision tree.

4. Repeat this procedure for all other remaining features and select the other features as internal nodes one by one. Repeat until the complete decision tree is formed.



Continuous Features

Question is

How does the decision tree work for continuous features???

Answer

Calculate Threshold Value for the continuous feature!!!

Age	Class
10	no
12	no
20	yes
30	no
35	yes
40	yes
60	yes

Threshold???

Threshold=20
Age<20 , Age>=20

Thank You…