AI & Robotics

Decision Trees

AI & ROBOTICS LAB

Goals

The **junior-colleague**

- can explain in their own words what a decision tree is
- can explain the difference between classification and regression in context of a decision tree
- can explain the usages of decision trees
- can explain how to decide on a split point for a classification tree based on the gini index
- can explain how to decide on a split point for a classification tree based on entropy and information gain
- can explain how to decide on a split point for a regression tree
- can construct a classification decision tree for a given problem based on the gini index
- can construct a classification decision tree for a given problem based on entropy and information gain
- can implement a decision tree for a given problem in *scikit-learn*
- can explain the advantages and disadvantages of decision trees

What is a decision tree?

- Predict the value of a target variable (y) by learning simple decision rules inferred from the data features (X)
- Each node represents a feature or attribute
- Each branch represents a decision or rule on that attribute
- Each leaf represents an outcome

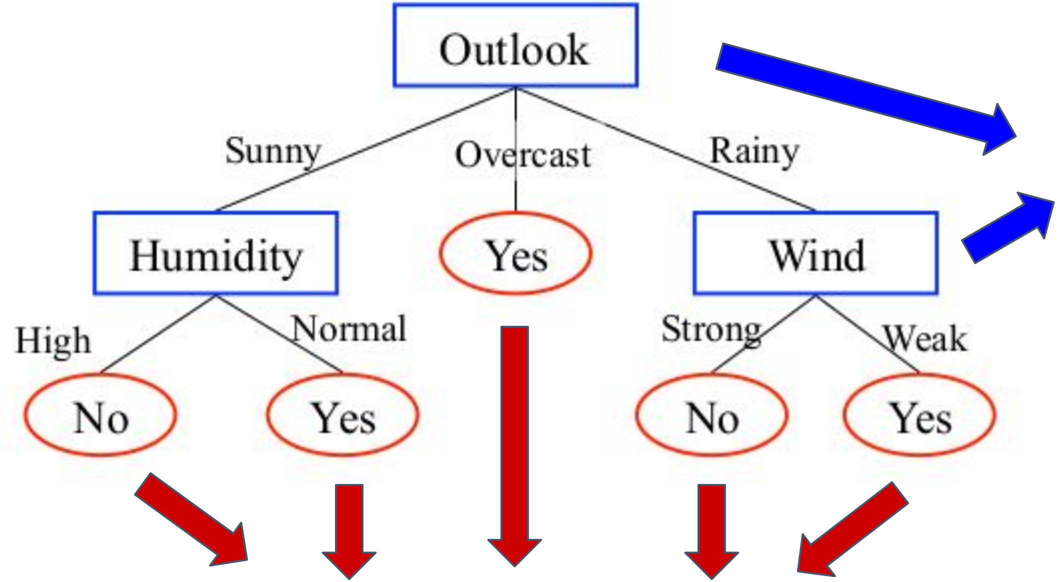
=> Used for classification and regression

=> Used for structured data

=> Nonlinear

What is a decision tree?

Are we gonna play tennis today?
Let's look at the weather!

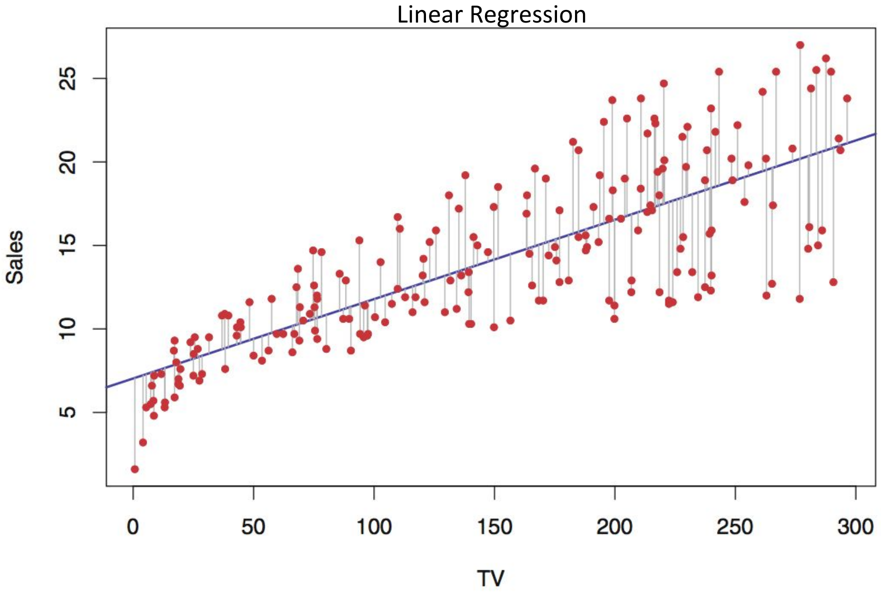


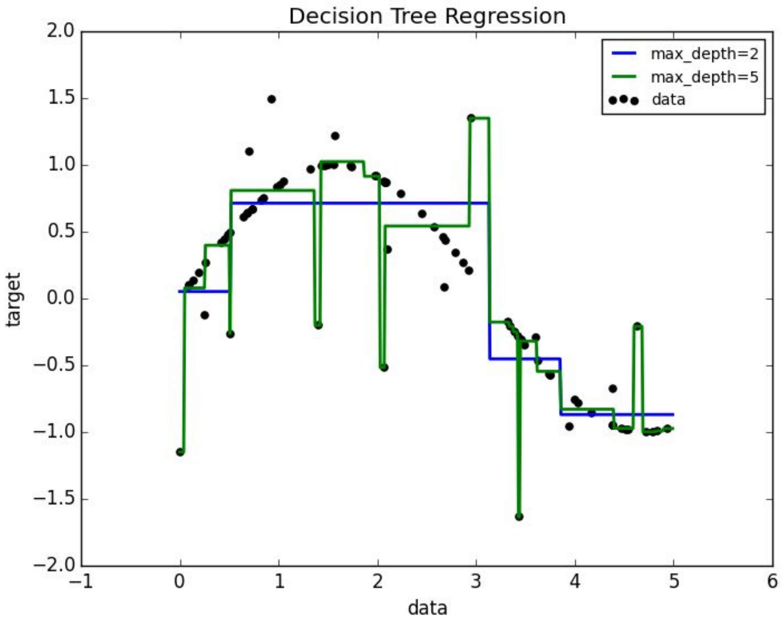
Internal nodes: condition on attributes/features

Leaf nodes: classification/prediction

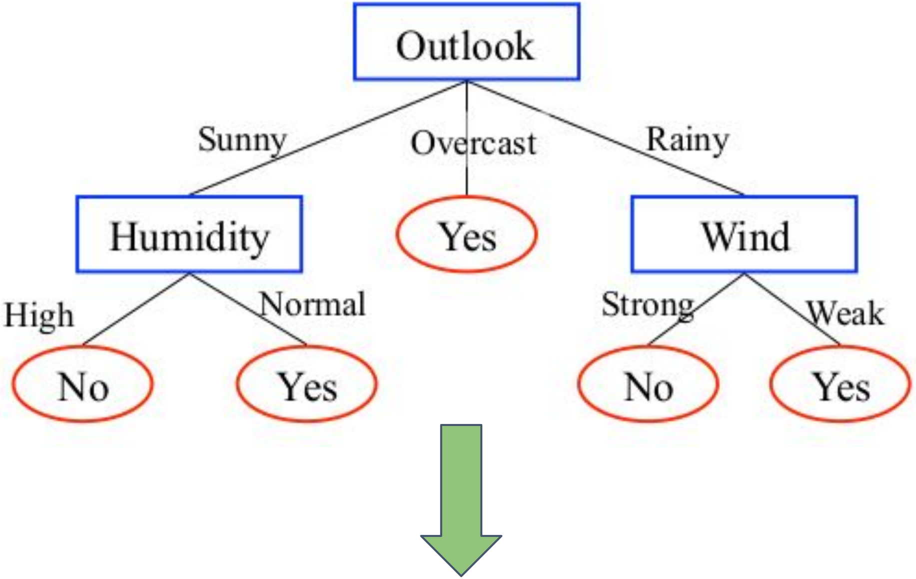
What is a decision tree?

It's nonlinear!



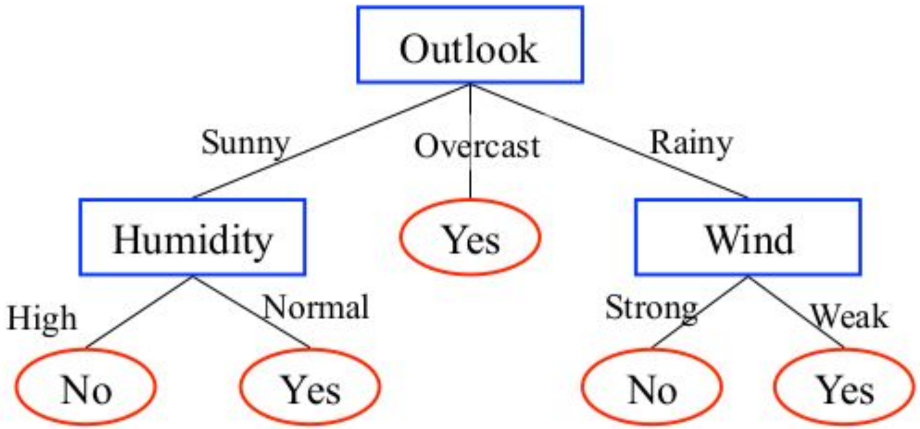


What is a decision tree?



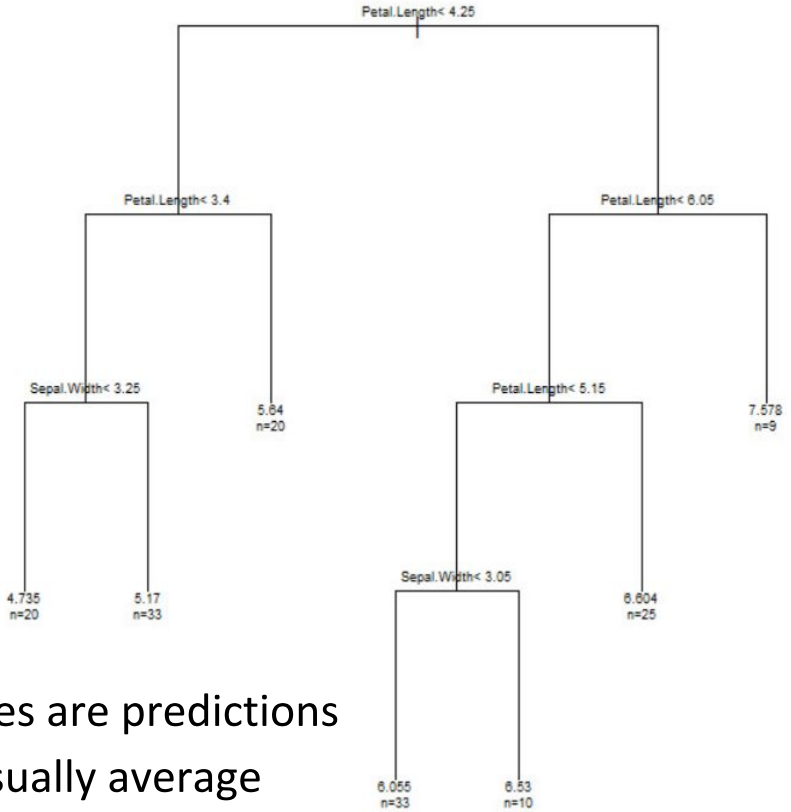
If Outlook = Sunny and Humidity = High then No
If Outlook = Sunny and Humidity = Normal then Yes

Classification trees



Leaf nodes are classes

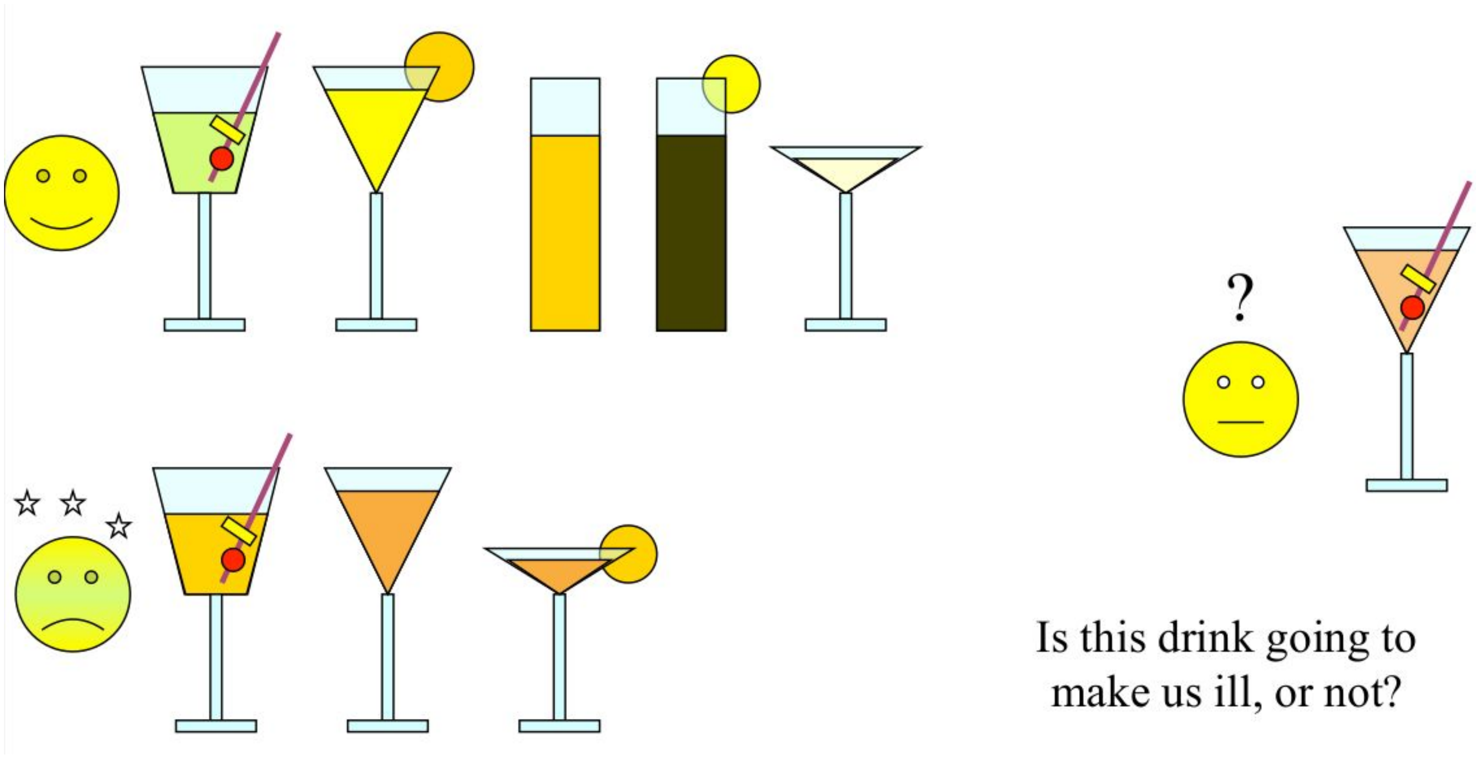
Regression trees



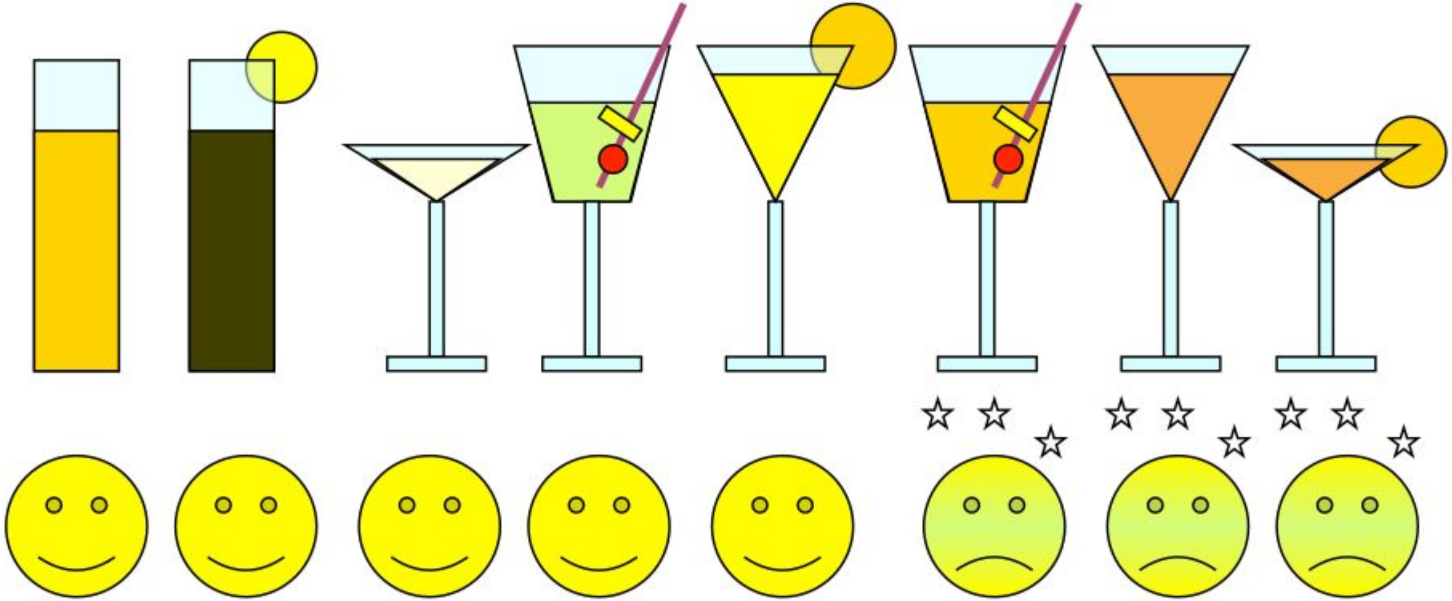
Leaf nodes are predictions

=> usually average

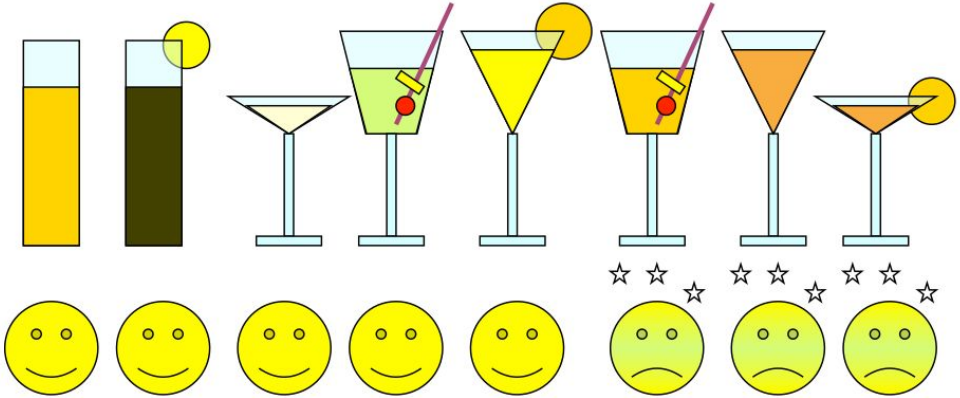
Example



Example

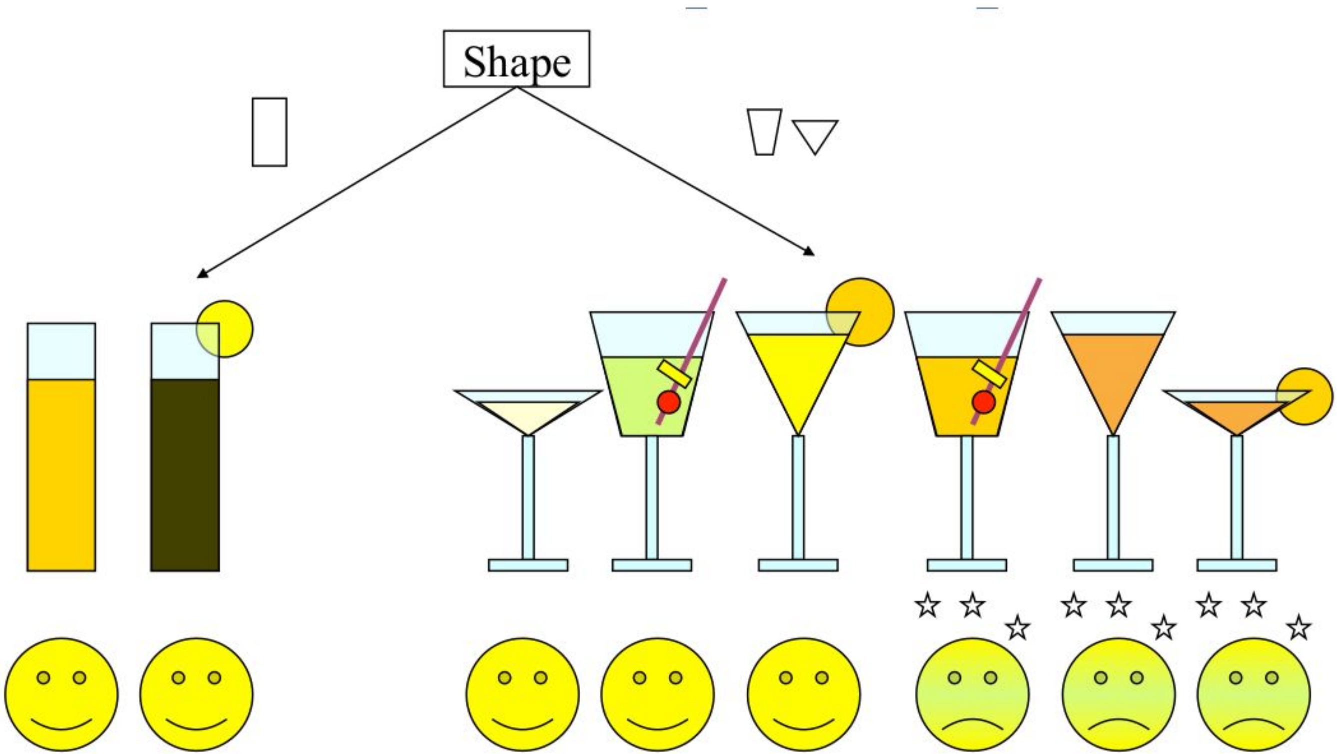


Example

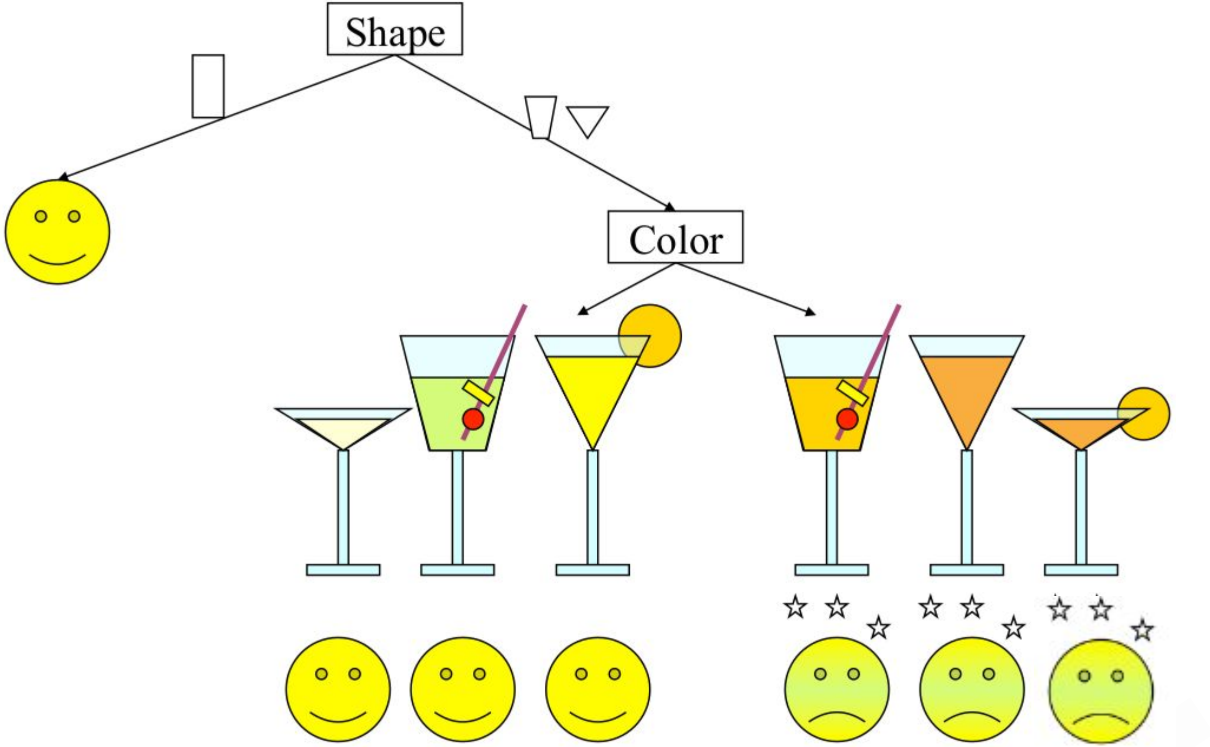


Shape	Color	Content	Sick?
Cilinder	Orange	25cl	No
Cilinder	Black	25cl	No
Coupe	White	10cl	No
Trapezoid	Green	15cl	No
Coupe	Yellow	15cl	No
Trapezoid	Orange	15cl	Yes
Coupe	Orange	15cl	Yes
Coupe	Orange	10cl	Yes

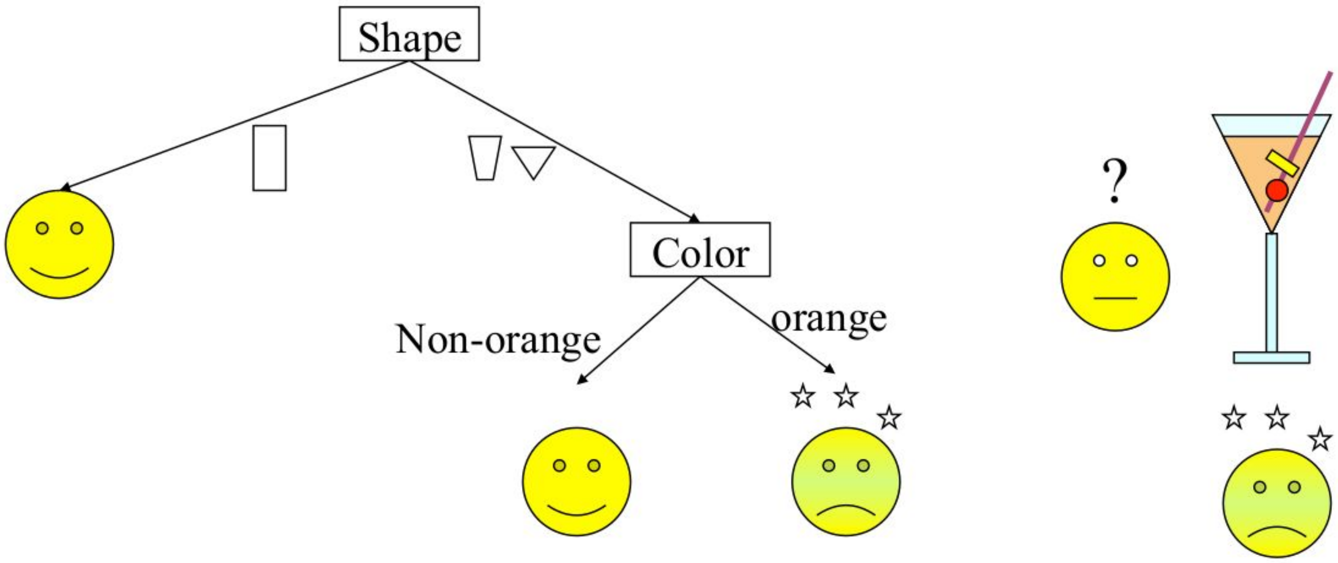
Example



Example



Example

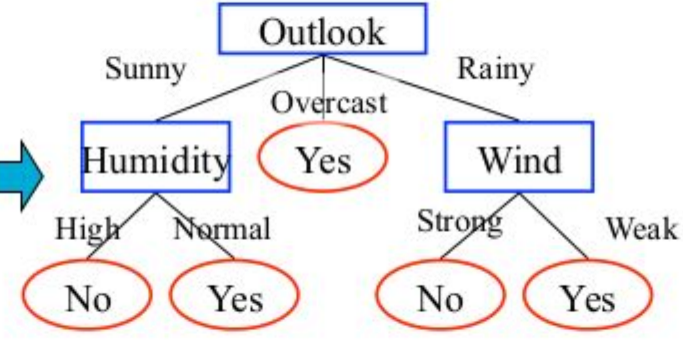


Basic algorithm

- Start with full data set
- Find test that partitions examples as well as possible

  => examples with same class, or otherwise similar examples, should be put together
- for each outcome of test, create child node
- Move examples to children according to outcome of test
- Repeat procedure for each child that is not “pure”, or until some stopping criterion is met

Outlook	Temp.	Hum.	Wind	Play?
Sunny	85	85	False	no
Sunny	80	90	True	no
Overcast	83	86	False	yes
...	...	...	...	...



How to decide?

- Uses heuristics at every decision node
- Try every split of every variable
- At each node decide:
  - Which variable to use
  - Which split point to use
- Metrics to decide on the prediction or classification at each leaf node:
  - Multiple possibilities
  - In practice: average works well

=> Different for classification and regression

How to decide - Classification

- Measure the homogeneity of the target variable within the subsets
- Metrics
  - Gini impurity
  - Information Gain

How to decide - Classification

Gini impurity

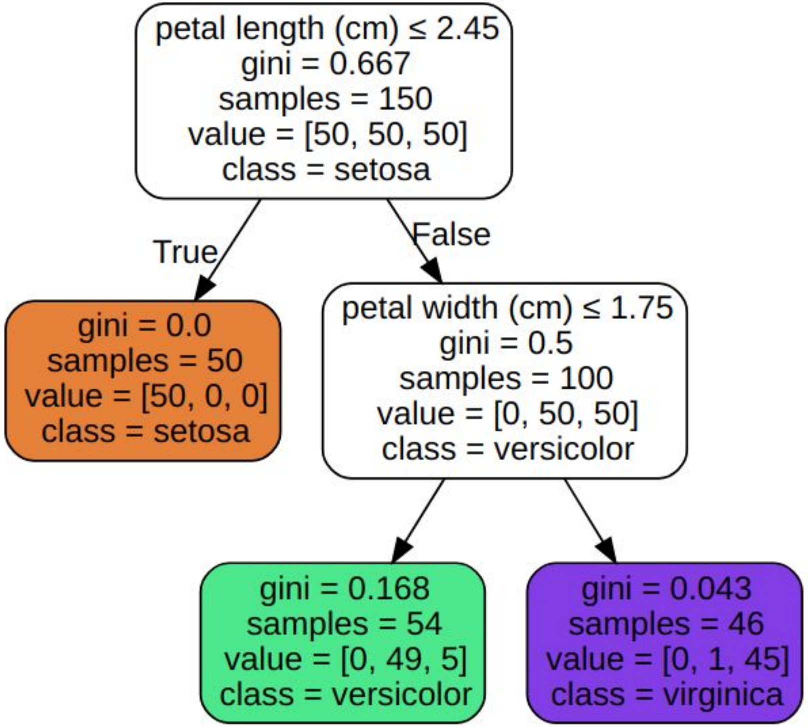
- Definition: a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset
- Reaches its minimum (0) when all cases in the node fall into a single target class
- Decision Tree algorithms will always try to minimize Gini impurity
- Formula:

$$\text{Gini}(S) = 1 - \sum p_i^2$$

How to decide - Classification

Gini impurity

- Example



$$\text{Gini}(S[\text{True}]) = 1 - \sum p_i^2 = 1 - (50/50)^2 = 0$$

$$\text{Gini}(S[\text{True}]) = 1 - \sum p_i^2 = 1 - (49/54)^2 - (5/54)^2 = 0.168$$

How to decide - Classification

Entropy

- Definition: Entropy is the measures of impurity, disorder or uncertainty in a bunch of examples
- Entropy controls how a Decision Tree decides to split the data
- Formula:

$$\text{Entropy}(S) = - \sum p_i \log_2(p_i)$$

How to decide - Classification

Information Gain

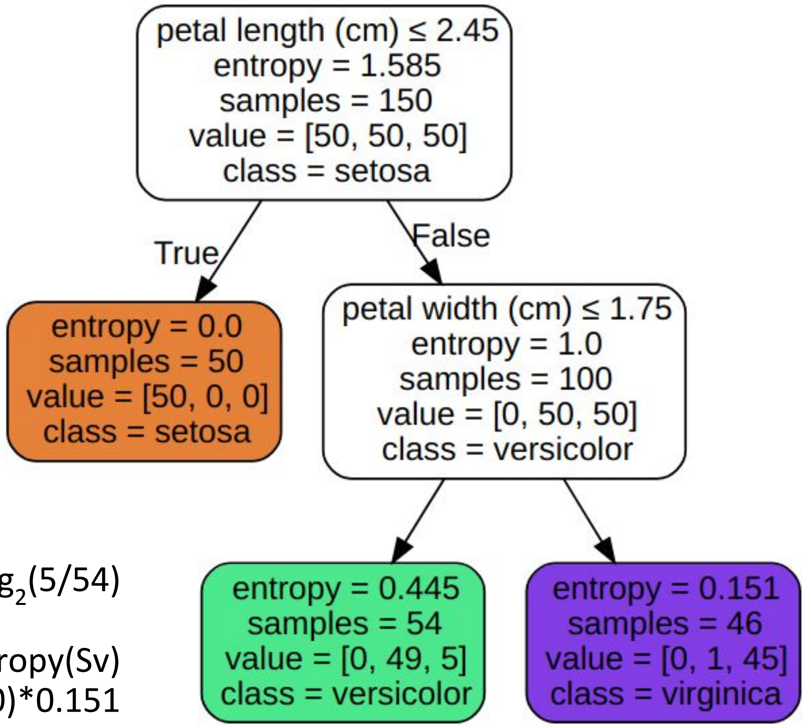
- Definition: Information gain (IG) measures how much “information” a feature gives us about the class
- Information gain is the main key that is used by Decision Tree Algorithms to construct a Decision Tree
- Decision Tree algorithms will always try to maximize Information gain
- The attribute with the highest Information gain will be split first
- Formula:

$$IG(S,Attr) = Entropy(S) - \sum |S_v| \,/\, |S| \; Entropy(S_v)$$

How to decide - Classification

Information Gain

- Example



$E(S[True]) = - \sum pi \log2(pi) = - (50/50)*\log_2(50/50) = 0$

$IG(S, \text{petal length}) = Entropy(S) - \sum |Sv| / |S| \ Entropy(Sv)$

$= 1.585 - (50/150)*0 - (100/150)*1$

$= 0.918$

$E(S[True]) = - \sum pi \log2(pi) = - (49/54)\log_2(49/54) - (5/54)\log_2(5/54)$

$= 0.445$

$IG(S, \text{petal width}) = Entropy(S) - \sum |Sv| / |S| \ Entropy(Sv)$

$= 1 - (54/100)*0.445 - (46/100)*0.151$

$= 0.69$

How to decide - Regression

- Mean Squared Error
- Minimize variance

See notebook for example

Why decision trees?

Advantages	**Disadvantages**
Simple to understand and to interpret (white box)	Do not generalise the data well (overfitting)
Easy to visualize	Vulnerable to small variations in the data
Able to handle both numerical and categorical data	Difficult to learn an optimal decision tree
Able to handle multi class classification	
Fast