Decision Trees

Week 06 - Day 05

Let’s play a game!

Guess the character

(real/fantasy, dead/alive)

Binary answers (yes/no)

Jorina

“Male or female?”

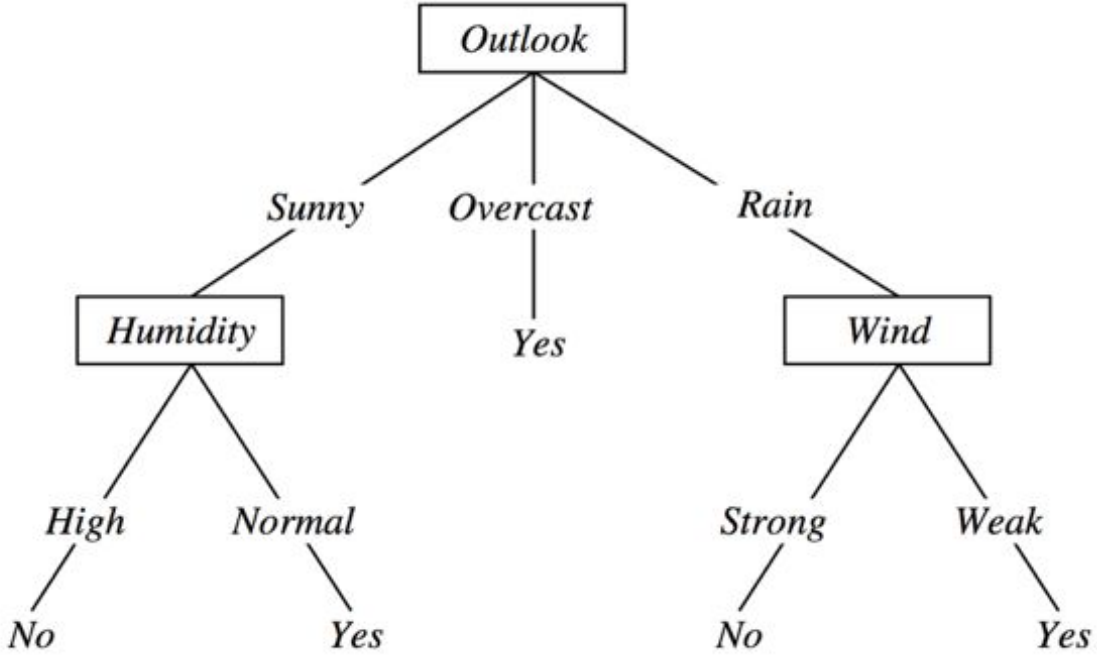
Vs.

“Was he/she one of the

5 first presidents of Spain?”

Decision Trees

Should I play golf?



Classification

+

Regression

How to build a decision tree

Basic principle 1:

Divide et impera

Basic principle 2:

Select the best node at every stage

(greedy approach)

https://www.youtube.com/watch?v=eKD5gxPPeY0

https://www.youtube.com/watch?v=LDRbO9a6XPU

Overfitting

Decision trees can overfit

Solution: pruning

http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier

Pros/cons

+ Non linear
+ White box!
+ No assumptions
+ They can manage categorical/numerical features

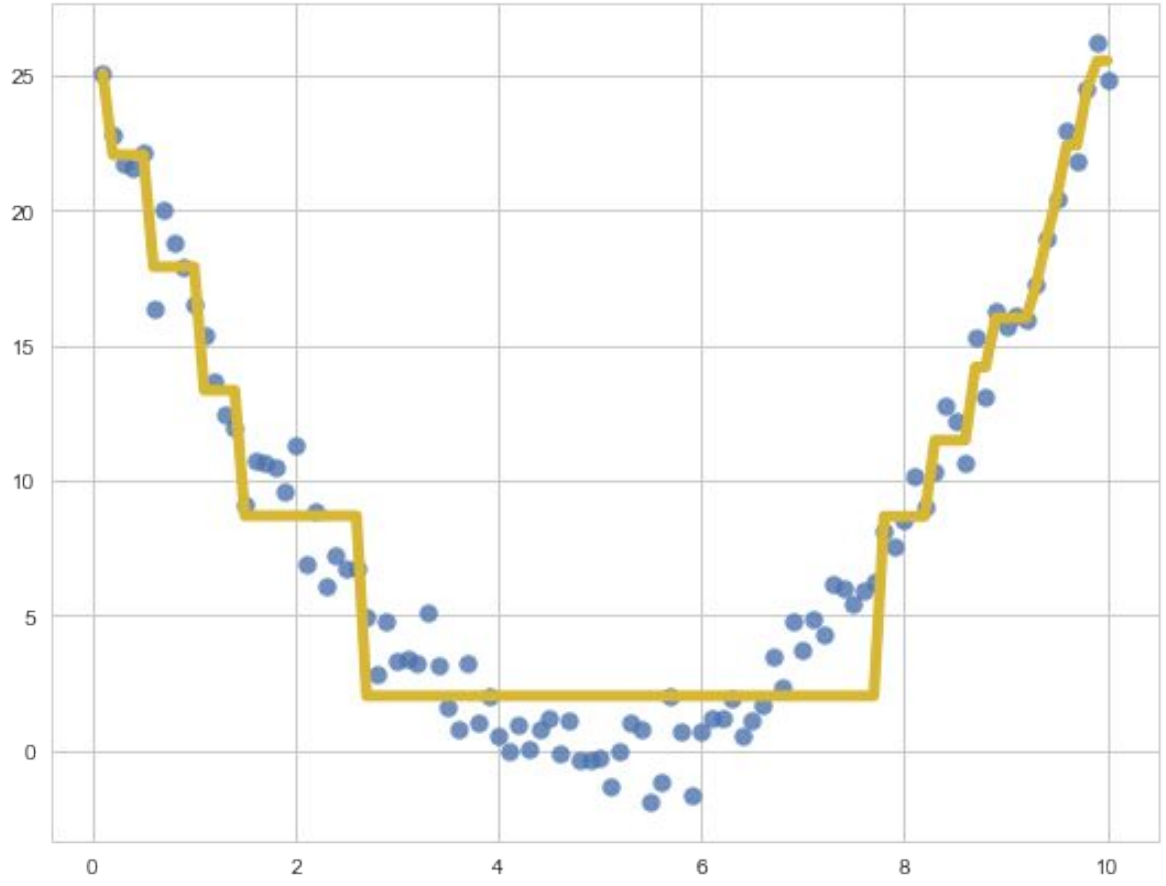
+ Multiclass
+ They can manage null values
+ Very good at describing the data

- Not very good performances
- They can overfit (but you can control it)
- They can have problems with unbalanced classes


They are aggregated together to create the best models in ML

Question 1

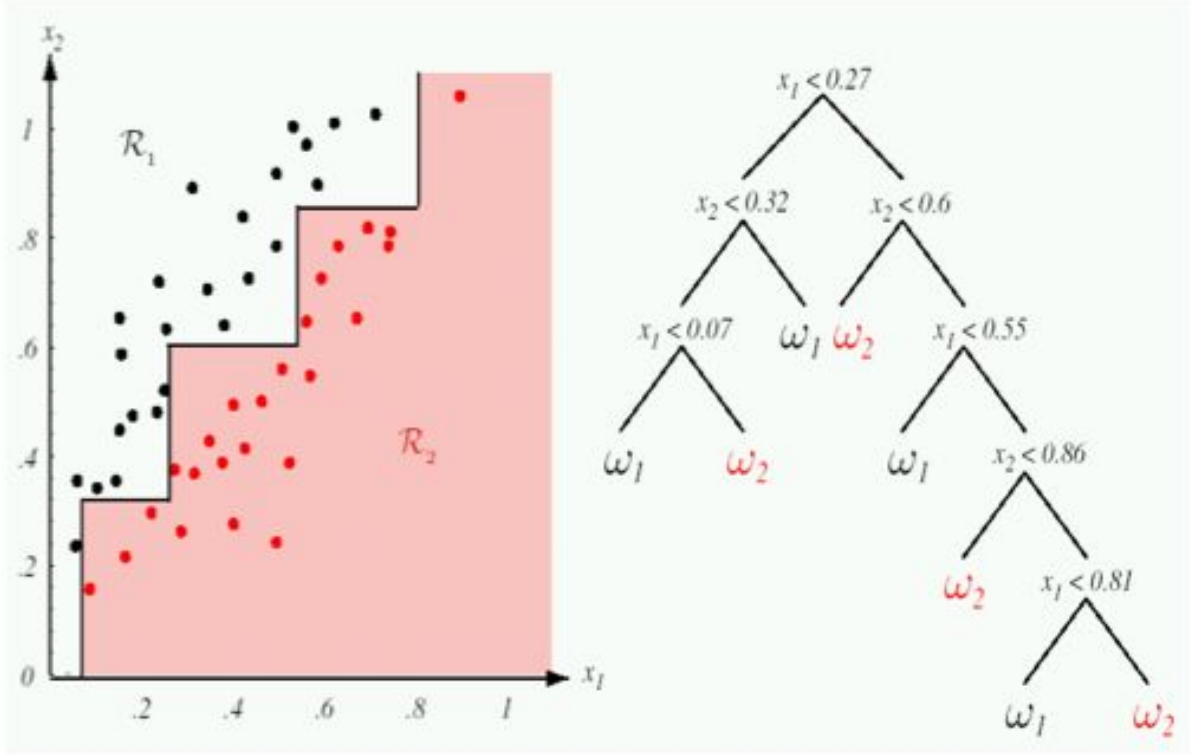
What does the decision boundary of a decision tree look like?


25
20
15
10
5
0
0
2
4
6
8
10

Question 2

When is a linear model

better than a tree?


x_2
1
.8
.6
.4
.2
0
$\mathcal{R}_1$
$\mathcal{R}_2$
.2
.4
.6
.8
1
x_1
$x_1 < 0.27$
$x_2 < 0.32$
$x_2 < 0.6$
$x_1 < 0.07$
ω_1
ω_2
$x_1 < 0.55$
ω_1
ω_2
ω_1
$x_2 < 0.86$
ω_2
$x_1 < 0.81$
ω_1
ω_2