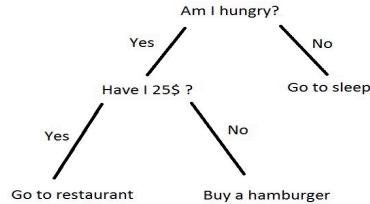


Decision Tree

- A very natural and intuitive model; used for both classification and regression

Week 9: Decision Tree & Random Forest



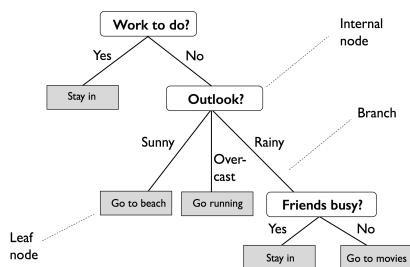
[Another example](#)

https://archive.nytimes.com/www.nytimes.com/imagepages/2008/04/16/us/20080416_OBAMA_GRAPHIC.html?emc=polb1&nl=pol

1

2

Decision Tree



- Internal nodes: tests on features
- Branches: feature attribute value; one branch for each value;
- Leaf nodes: assign classification / output

3

Day	Temperature	Outlook	Humidity	Windy	Play Golf?
07-05	hot	sunny	high	false	no
07-06	hot	sunny	high	true	no
07-07	hot	overcast	high	false	yes
07-09	cool	rain	normal	false	yes
07-10	cool	overcast	normal	true	yes
07-12	mild	sunny	high	false	no
07-14	cool	sunny	normal	false	yes
07-15	mild	rain	normal	false	yes
07-20	mild	sunny	normal	true	yes
07-21	overcast	high	true	yes	yes
07-22	hot	overcast	normal	false	yes
07-23	mild	rain	high	true	no
07-26	cool	rain	normal	true	no
07-30	mild	rain	high	false	yes

today	cool	sunny	normal	false	?
tomorrow	mild	sunny	normal	false	?

How to Build a decision tree? Which node is the root and how to to split the branches?

Decision Tree Learning

- Decision trees can express any Boolean function.
- Goal: Finding a decision tree that agrees with training set.

How about construct a decision tree that has one path to a leaf for each example, where the path tests sets each attribute value to the value of the example?

Choosing the Best Split

- Ockham's Razor: All other things being equal, choose the simplest explanation
- Decision Tree Induction:
 - Find the smallest tree that classifies the training data correctly
 - Good balance between bias v.s. variance
- Approach
 - Use heuristic search (a **greedy** search)

Basic algorithm

- Tree is constructed in a top-down recursive manner
- Test attributes are selected on the basis of a heuristics or statistical measure (e.g., information gain)
- Examples are partitioned recursively based on selected attributes

7

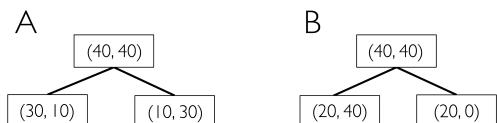
Building a Decision Tree

- Which split is better?

The goal is to reduction in uncertainty in each split



8



8

Measure of Information Gain

- Need a quantitative measure:
 - Gini impurity
 - Entropy
 - Classification errors
- Gini impurity: the probability of obtaining **two different outputs**.

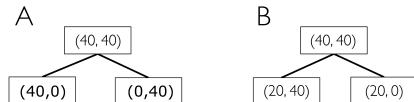
$$gini(D) = 1 - \sum_{j=1}^n p_j^2$$

9

Measure of Information Gain

- Information gain can be calculated as:

$$\text{Gini}(\text{ParentNode}) - \text{Sum of } \#(\text{ChildNode})/\#(\text{ParentNode}) * \text{Gini}(\text{ChildNode})]$$
- Higher Information Gain = Better Split



10



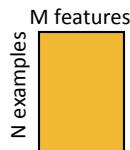
Random Forest Classifier

- Use a bagging of un-pruned decision trees with a randomized selection of features at each split.
- An ensemble classifier to reduce variance.

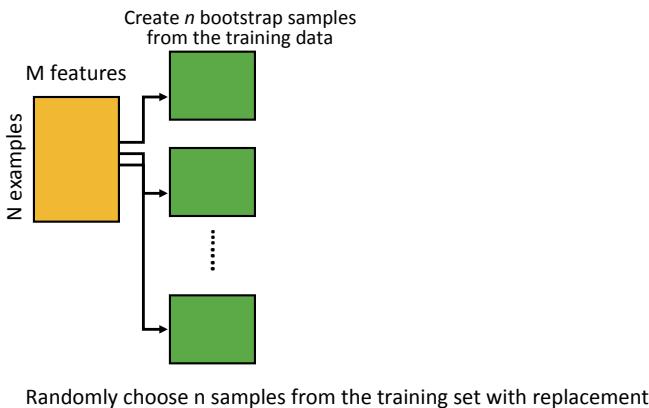
11

Random Forest Classifier

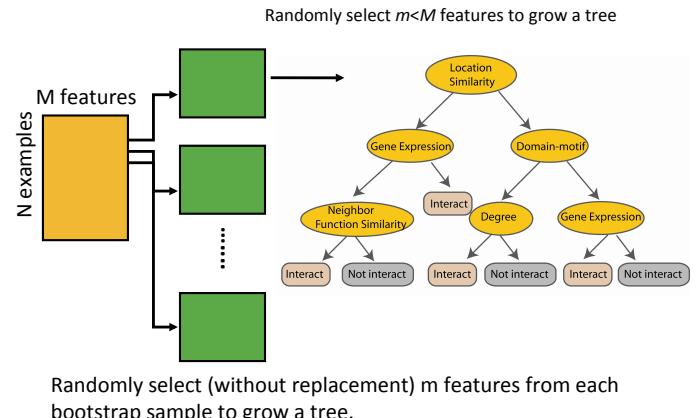
Training Set: N examples with M features



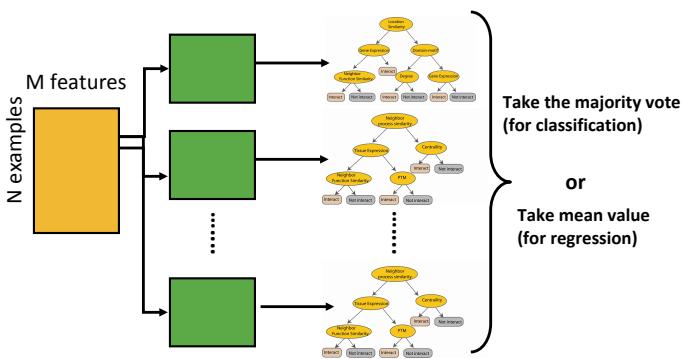
Random Forest Classifier



Random Forest Classifier



Random Forest Classifier



Random Forest v.s. SVM

- Has similar performance to SVM
- Fewer hyper-parameters
- Help provide insights into the data
- Faster to train and predict
- Fully parallelizable – even faster

16

Question

Are Decision Tree and Random Forest classifier sensitive to feature scaling ?