Pattern Recognition

Yukyung Choi

yk.choi@rcv.sejong.ac.kr

Summary

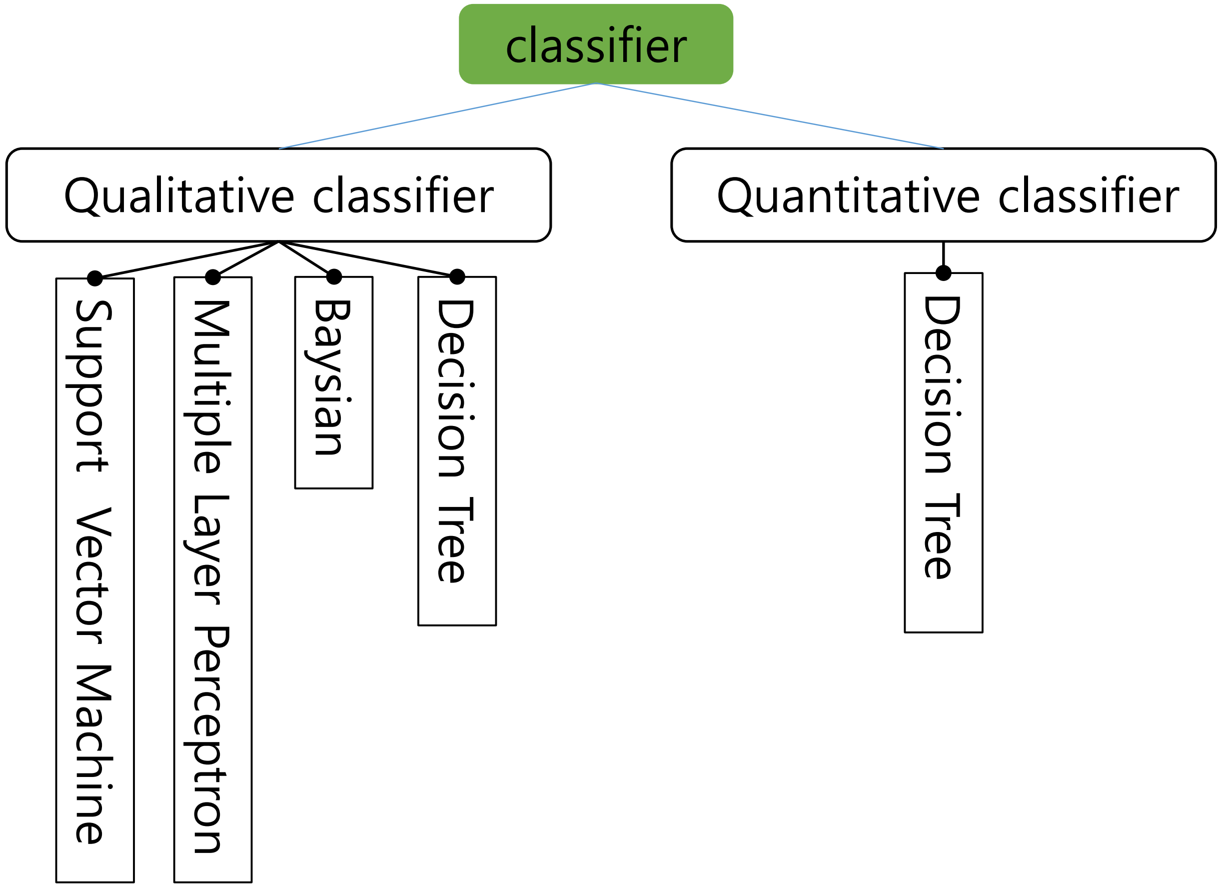
- Decision Tree & Random Forests I
  - Hand on Labs: Data Classification using library

- Decision Tree & Random Forests II
  - Hand on Labs: Implementation of DT-RF

Content

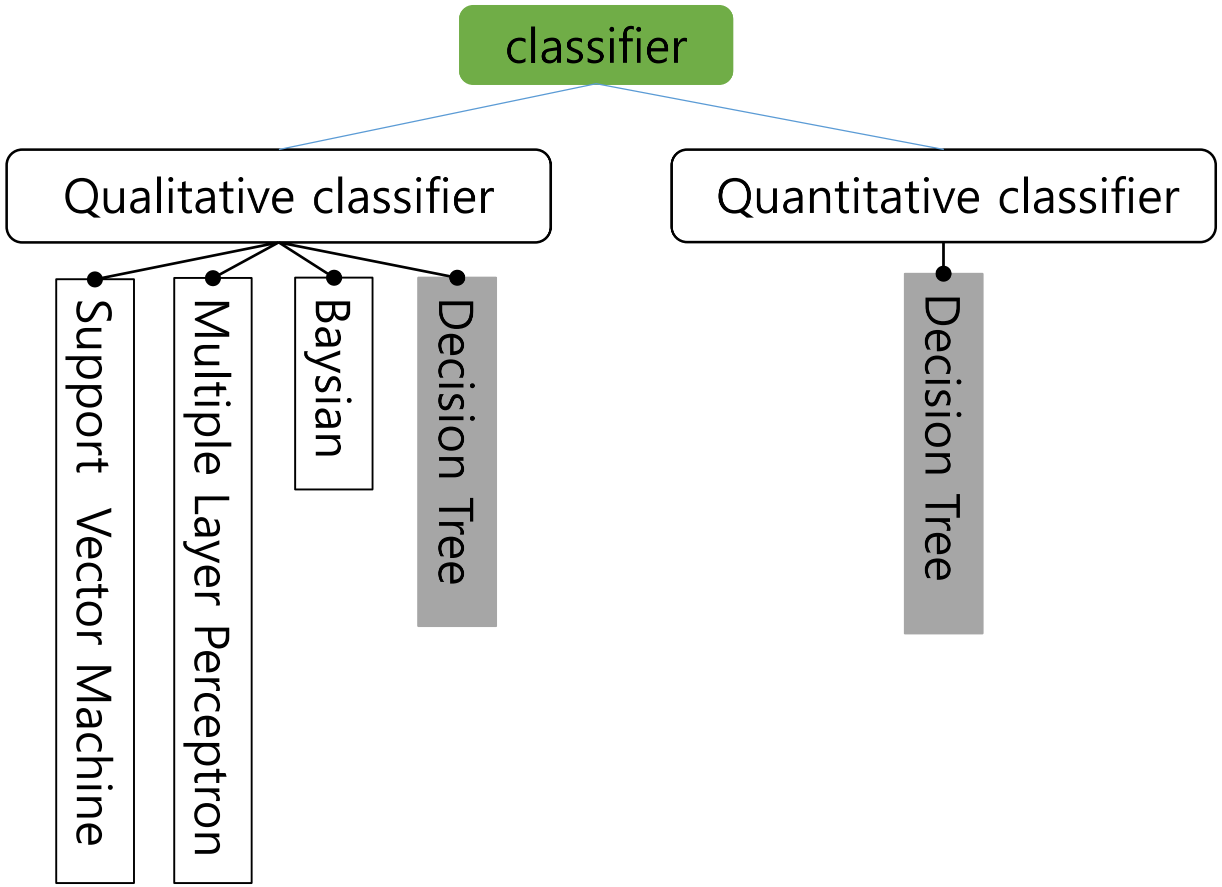
- Decision Tree
- Random Decision Trees
- Random Forests
- Random Forests with discriminative decision tree

Classifier



Metric/Non-Metric data : 어떤 사물이나 개념을 양적/질적 으로 표시한 데이터

Classifier



Metric/Non-Metric data : 어떤 사물이나 개념을 양적/질적 으로 표시한 데이터

Decision Tree

Definition

Tree structure based Classifier

Classification method by using several rules and constrains.

tree [1] : a tree is a widely-used data structure that emulates a hierarchical tree structure with a set of linked nodes.

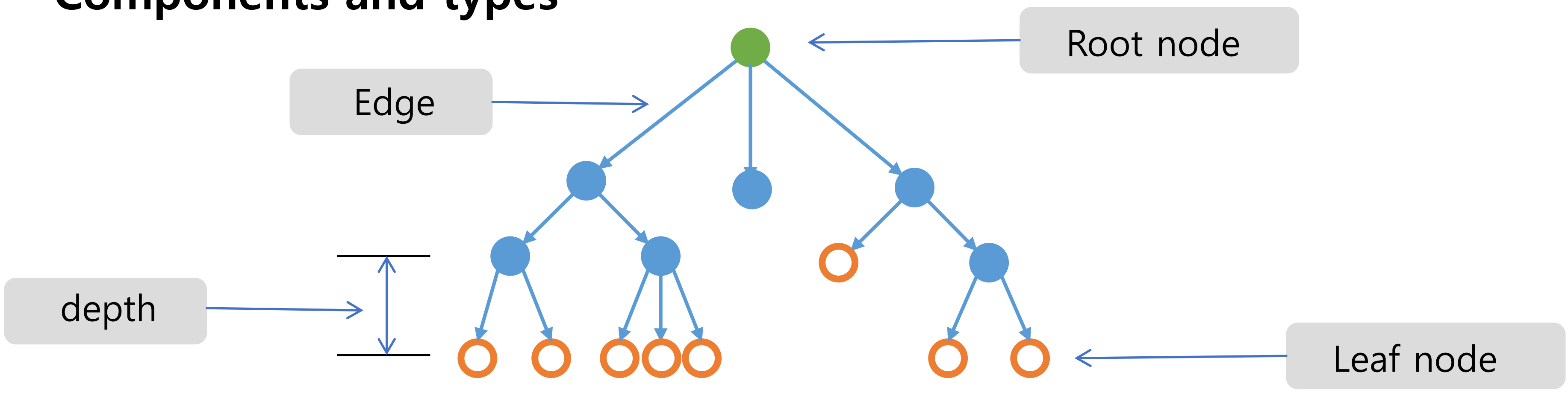
Decision Tree

Definition

Tree structure based Classifier

Classification method by using several rules and constrains.

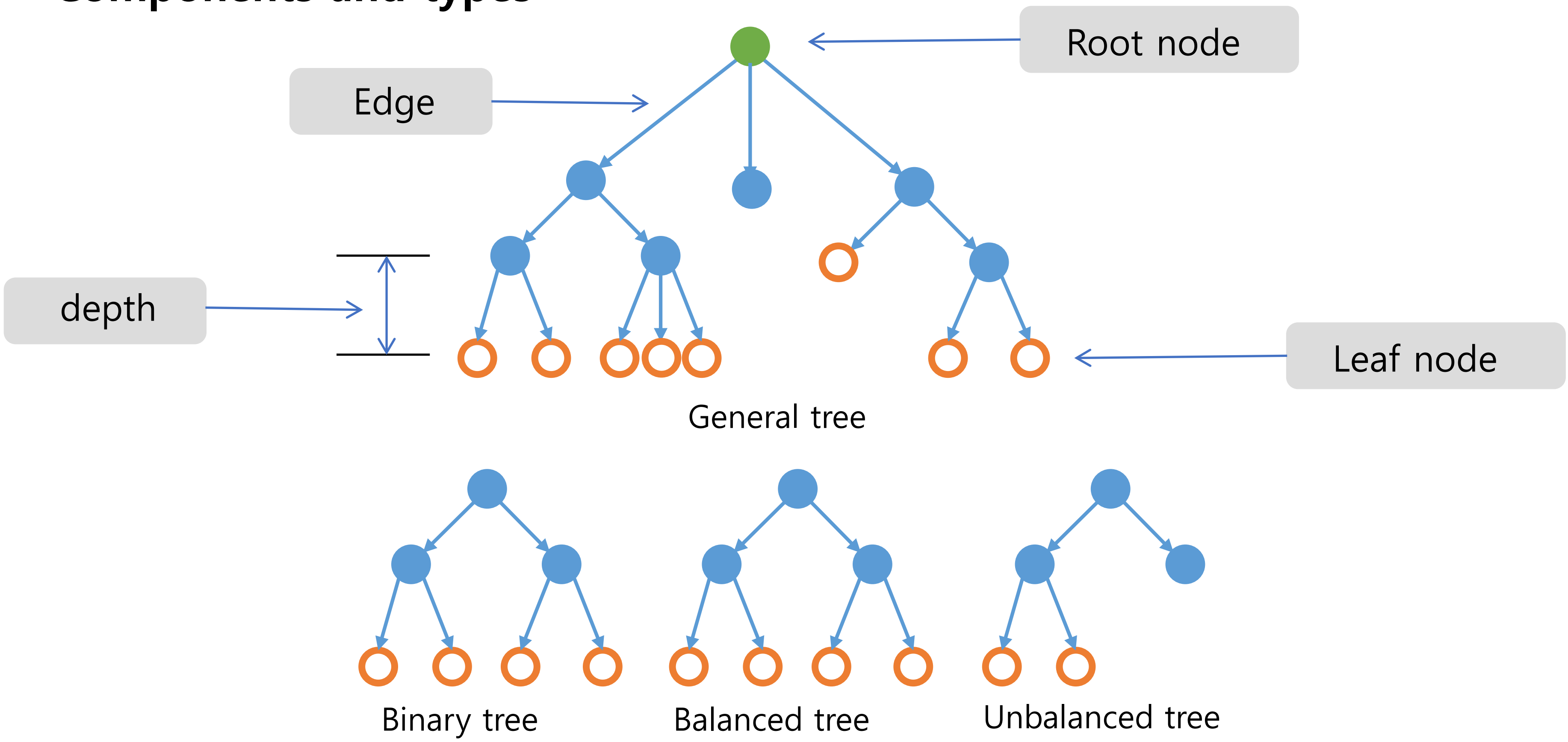
Components and types



General tree

Decision Tree

Components and types



Decision Tree

Training

build the tree

Consideration in this step

1) How many splits should there be at each node?
2) How to select the query attribute?
3) When to stop growing the tree?
4) Which class to allow leaf nodes?

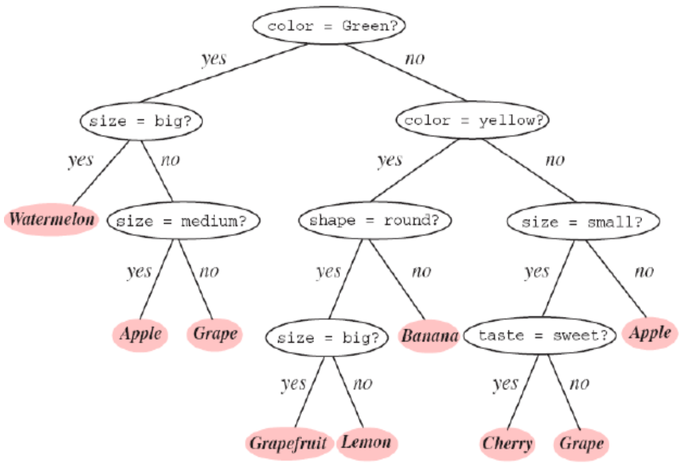
Decision Tree

Training

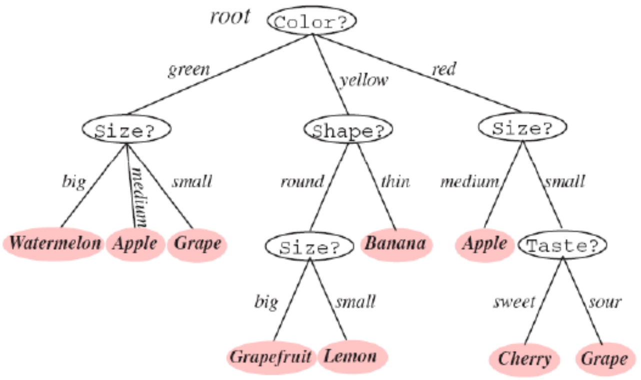
build the tree

Consideration in this step

1) How many splits should there be at each node?



Multi-valued tree



Binary tree

Decision Tree

Training

build the tree

Consideration in this step

2) How to select the query attribute?

Ready to Candidate query attributes :
Collect all possible attributes

Decision Tree

Training

build the tree

Consideration in this step

2) How to select the query attribute?

Ready to Candidate query attributes :
Collect all possible attributes

Select the best query attributes :
Find the maximal decrease in *impurity(entropy)*

Decision Tree

Training

build the tree

Consideration in this step

2) How to select the query attribute?

Ready to Candidate query attributes :
Collect all possible attributes

Select the best query attributes :
Find the maximal decrease in *impurity(entropy)*

$$\Delta im(T) = im(T) - \frac{|X_{Tleft}|}{|X_T|} im(T_{left}) - \frac{|X_{Tright}|}{|X_T|} im(T_{right})$$

$$\Delta im(T) = \frac{|X_{Tleft}|}{|X_T|} \frac{|X_{Tright}|}{|X_T|} \left(\sum_{i=1}^{M} | p(w_i | T_{left}) - p(w_i | T_{right}) | \right)^2$$

Decision Tree

Training

build the tree

Consideration in this step

2) How to select the query attribute?

Impurity functions

Entropy, Gini, Misclassification

Entropy	$im(T) = 1 - \sum_{i=1}^{M} p(w_i \mid T)^2 = \sum_{i \neq j} p(w_i \mid T) p(w_j \mid T)$
Gini	$im(T) = -\sum_{i=1}^{M} p(w_i \mid T) \log_2 p(w_i \mid T)$
Misclassification	$im(T) = 1 - \max_i p(w_i \mid T)$

Decision Tree

Training

build the tree

Consideration in this step

3) When to stop growing the tree?

Stop conditions

$im(T) = 0$ → **Overfitting**

Decision Tree

Training

build the tree

Consideration in this step

3) When to stop growing the tree?

Stop conditions

$im(T) = 0$ → **Overfitting**

the number of X_t <= threshold

argmax $\triangle im(T)$ <= threshold → **Premature convergence**

Decision Tree

Training

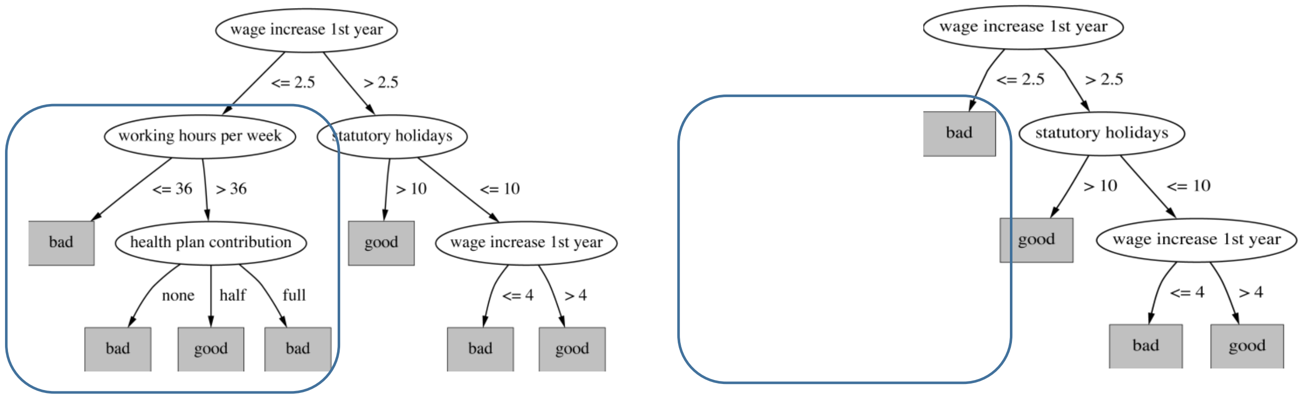
build the tree

Consideration in this step

3) When to stop growing the tree?

Avoiding overfitting & premature convergence

Make the largest tree and then do *pruning*



Example of pruning

Decision Tree

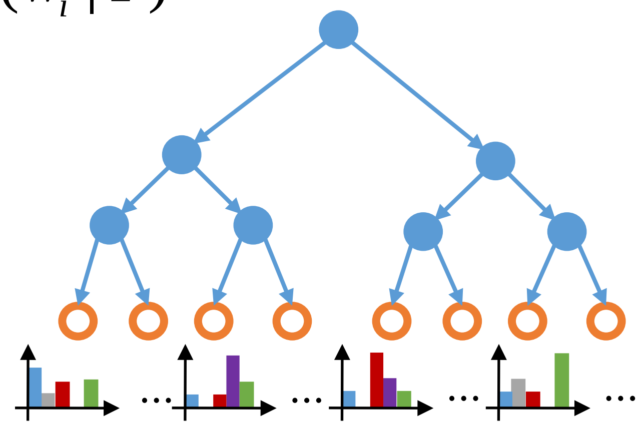
Training

build the tree

Consideration in this step

4) Which class to allocate leaf nodes?

$$K = \arg\max_{i} p(w_i \mid T)$$



Decision **T**ree

Classification

recognize the sample

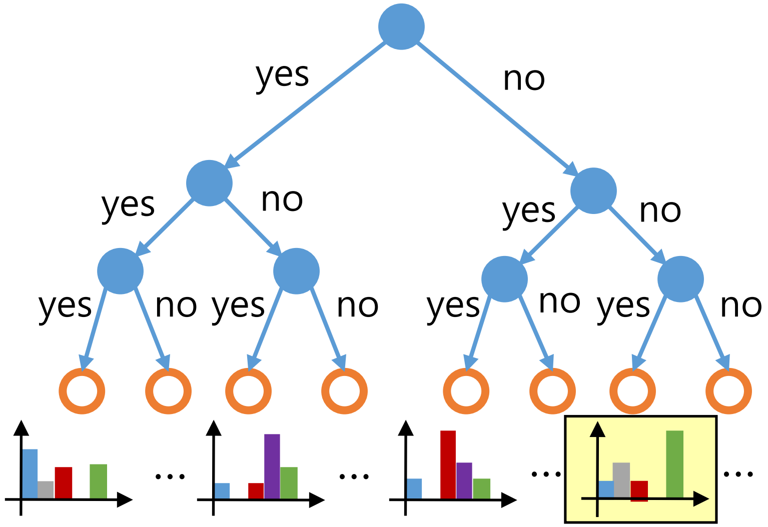
Decision Tree

Classification

recognize the sample

Advantages

Very fast "Yes/NO" operation
Just h-1 times comparison (h : depth level)



Example of classification

Decision Tree

Characteristics

+Handling the metric and non-metric data

weight

height

distance

cost

Metric data

gender

Blood type

job

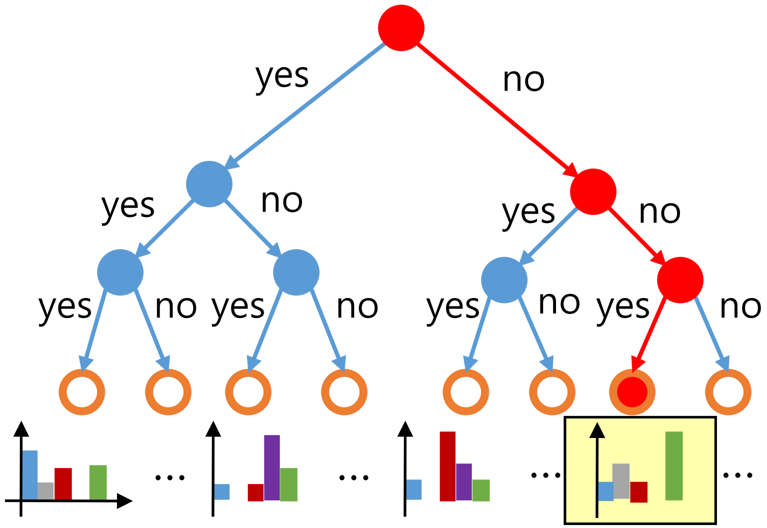
Brand name

Non-Metric data

Decision Tree

Characteristics

+Handling the metric and non-metric data
+Visualization



Decision **T**ree

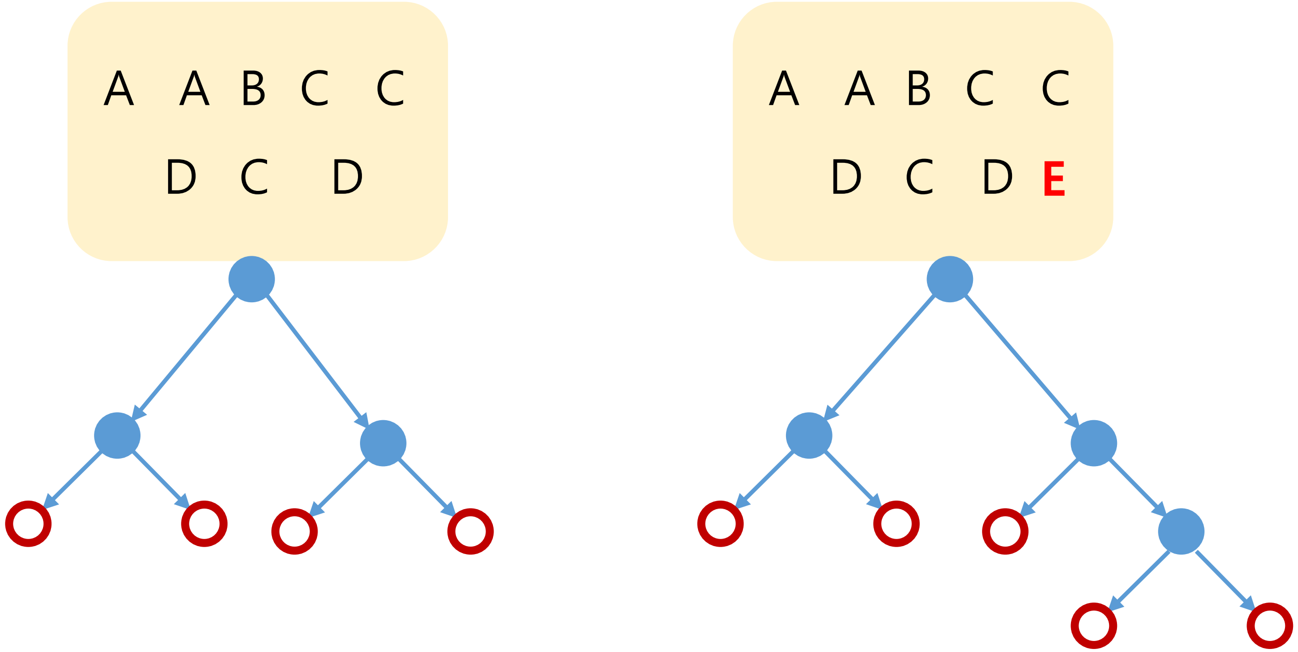
Characteristics

+Handling the metric and non-metric data
+Visualization
+Fast recognition

Decision Tree

Characteristics

+Handling the metric and non-metric data
+Visualization
+Fast recognition
-Instability



Decision Tree

Characteristics

+Handling the metric and non-metric data
+Visualization
+Fast recognition
-Instability
-Greedy algorithms

Decision Tree

Characteristics

+Handling the metric and non-metric data
+Visualization
+Fast recognition
-Instability
-Greedy algorithms
+Handling the missing data

Decision Tree

Algorithms

CART, ID3, C4.5

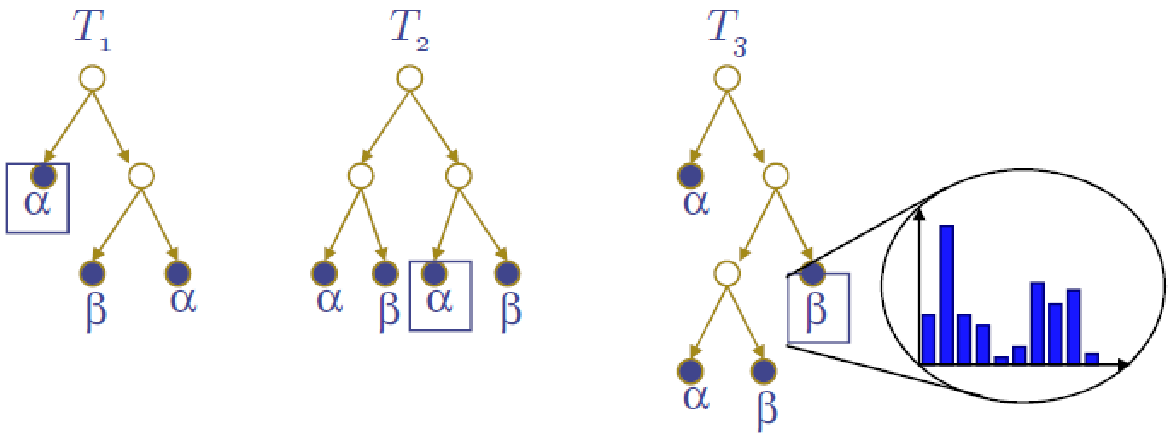
L.Breiman	R.Quinlan
CART	ID3
	↓ **extension**
	C4.5
	↓ **commercialization**
	C5.0

Comparisons

Property	CART	ID3	C4.5
Float data	O	X	O
Tree type	Binary	Multi	Multi
Prune	O	X	O
Classification	O	O	O
Regression	O	X	X
Missing data	Surrogate split	X	skip
Multi variable	O	X	X

Randomized Decision Trees

Randomized Decision Trees (Amit & German 1997)
Multiple classifier of several trees



Randomized decision trees

Randomized **D**ecision **T**rees

Randomized Decision Trees (Amit & German 1997)
Multiple classifier of several trees

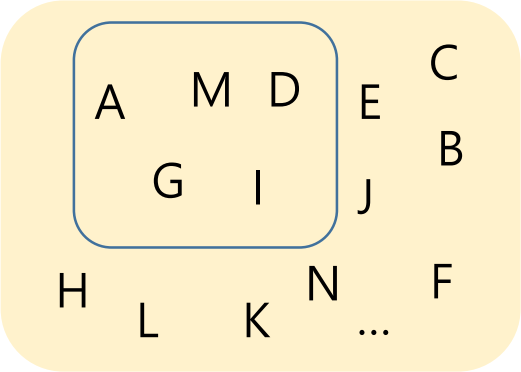
Idea : randomized attribute selection

Randomized Decision Trees

Randomized Decision Trees (Amit & German 1997)
Multiple classifier of several trees

Idea : randomized attribute selection

Instead randomly use subset of K attributes (K << All cases)
- Reduce correlation between different trees
- Typical choice : K = 10 for root node, K = 100*d (d : depth level)



All cases of possible attributes

Randomized Decision Trees

Randomized Decision Trees (Amit & German 1997)
Multiple classifier of several trees

Idea : randomized attribute selection

Instead randomly use subset of K attributes (K << All cases)
- Typical choice : K = 10 for root node, K = 100*d (d : depth level)
- Reduce correlation between different trees.

Choose best splitting attribute
- Minimizing entropy (impurity)

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

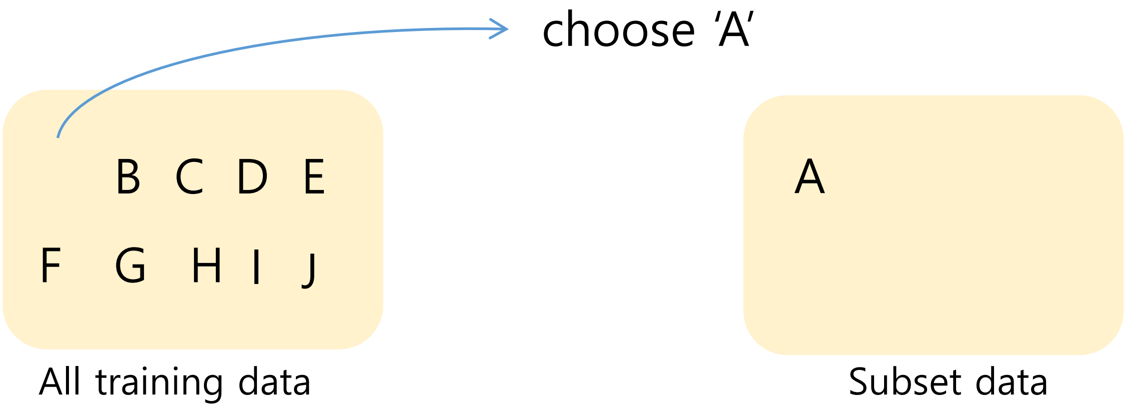
Idea : Randomness = 1)Bootstrap sampling + 2)randomized attribute selection

Randomized **F**orests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = [1]Bootstrap sampling + [2]randomized attribute selection

Bootstrap sampling
Select a subset by choosing N times with replacement from all training data.

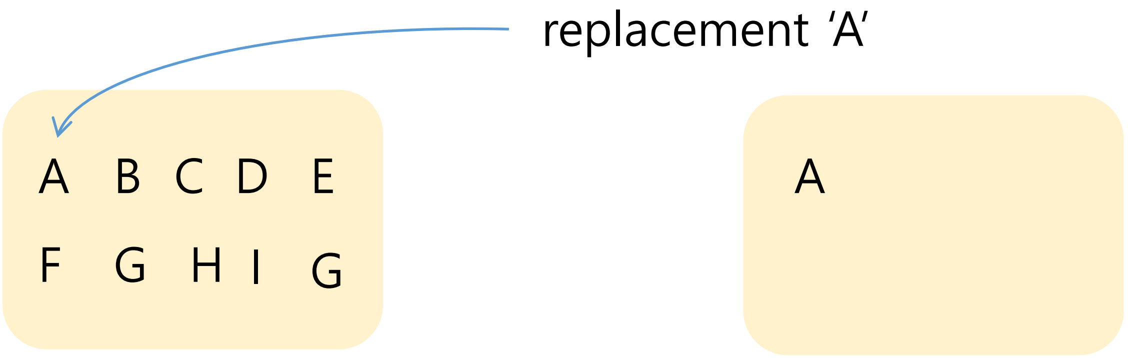


Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = 1)Bootstrap sampling + 2)randomized attribute selection

Bootstrap sampling
Select a subset by choosing N times with replacement from all training data.



All training data

Subset data

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = 1)Bootstrap sampling + 2)randomized attribute selection

Bootstrap sampling
Select a subset by choosing N times with replacement from all training data.

Randomized attribute selection
Instead randomly use subset of K attributes
- Typical choice : K = $\sqrt{N}$ **(** N : the number of subset)

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = 1)Bootstrap sampling + 2)randomized attribute selection

Advantage
Resistant to Overfitting

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = [1]Bootstrap sampling + [2]randomized attribute selection

Advantage

Resistant to Overfitting
Well suited for large training data

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = 1)Bootstrap sampling + 2)randomized attribute selection

Advantage

Resistant to Overfitting
Well suited for large training data
Empirically very good results. (≥ SVM, ≥ Boosting)

Randomized Forests

Randomized Forests (Breiman 2001)
Multiple classifier of several trees

Idea : Randomness = [1]Bootstrap sampling + [2]randomized attribute selection

Advantage
Resistant to Overfitting
Well suited for problems with large training data
Empirically very good results. ($\geq$ SVM, $\geq$ Boosting)

Disadvantage
Memory consumption

Comparison with various classifier

- CIFAR-10: Image Classification
- Comparison with various classifier
  - https://github.com/PhilippeCodes/Image-Classifier
  - https://github.com/PhilippeCodes/Image-Classifier/blob/master/Decision%20trees%20and%20random%20forests.ipynb

	Estimator	Test Accuracy
0	Baseline (dummy)	0.222
1	KNeighbors	0.776
2	DecisionTree	0.646
3	RandomForest	0.800
4	LogisticRegression	0.840
5	SVM Linear Kernel	0.817
6	SVM RBF Kernel	0.823
7	Multilayer Neural Network	0.821
8	Convolutional Neural Network	0.777

Hands on Lab

- Decision Tree from scratch
- Random Forests from scratch

Application

- Titanic: Machine Learning from Disaster
- Solution: Random Forests
  - https://www.kaggle.com/mukultiwari/titanic-top-14-with-random-forest

Application

- Bike Sharing Demand
- Solution: Random Forests
    - https://www.kaggle.com/kwonyoung234/for-beginner