Machine Learning

- statistical methods
- input : observations $\mathbf{x}$
- goal : learn the function $f(\mathbf{x}) \mapsto \mathbf{y}$
- output : $\hat{\mathbf{y}} \approx \mathbf{y}$ minimizing prediciton errors on $\mathbf{x}$

Machine Learning

- statistical methods
- input : observations $\mathbf{x}$
- goal : learn the function $f(\mathbf{x}) \mapsto \mathbf{y}$
- output : $\hat{\mathbf{y}} \approx \mathbf{y}$ minimizing prediciton errors on $\mathbf{x}$

Two big families

- supervised : $\mathbf{y}$ given with $\mathbf{x}$ when learning
- unsupervised : $\mathbf{y}$ hidden when learning

Machine Learning

- statistical methods
- input : observations $\mathbf{x}$
- goal : learn the function $f(\mathbf{x}) \mapsto \mathbf{y}$
- output : $\hat{\mathbf{y}} \approx \mathbf{y}$ minimizing prediciton errors on $\mathbf{x}$

Two big families

- supervised : $\mathbf{y}$ given with $\mathbf{x}$ when learning
- unsupervised : $\mathbf{y}$ hidden when learning

Tons of algorithms

- Decision Trees, Random Forests
- Support Vector Machines
- Neural Networks

How to learn ?

- Collect $\mathbf{X}$s $\rightarrow$ $\mathbf{x}_{tr}$, $\mathbf{x}_{ts}$
- Compute a frontier to separate $\mathbf{x}_{tr}$ (according to associate $\mathbf{y}_{tr}$ or based on similarity)
- Evaluate the number of errors
- Repeat to reduce the number of errors

How to learn ?

- Collect **X**s $\rightarrow$ $\mathbf{x}_{tr}$, $\mathbf{x}_{ts}$
- Compute a frontier to separate $\mathbf{x}_{tr}$ (according to associate $\mathbf{y}_{tr}$ or based on similarity)
- Evaluate the number of errors
- Repeat to reduce the number of errors

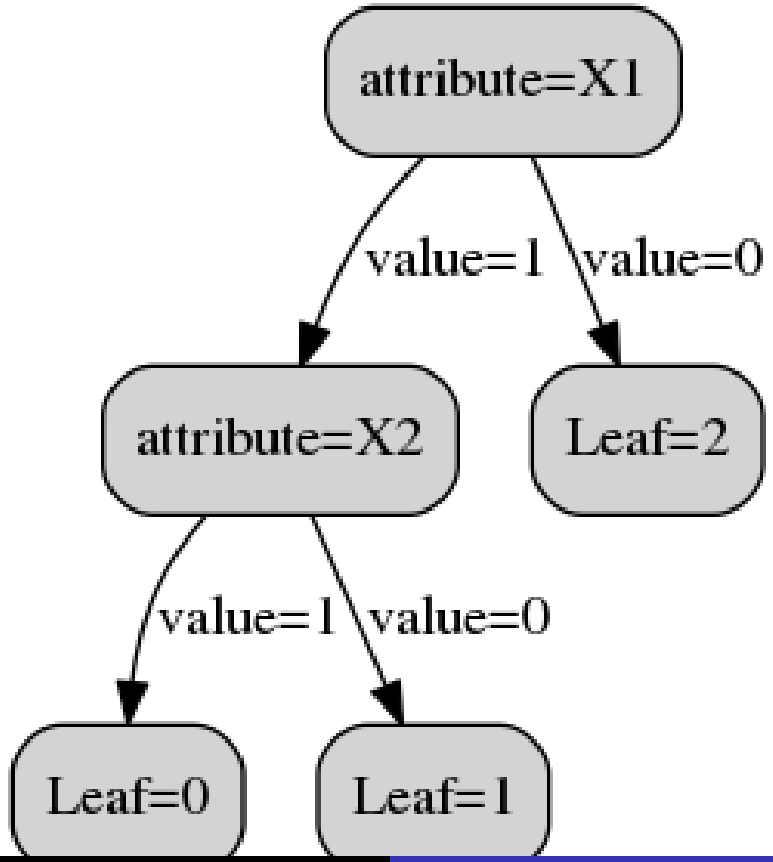
Overfitting

- If too much repetitions...
- No errors $\mathbf{x}_{tr}$
- But what about $\mathbf{x}_{ts}$ (unseen in the training phase)

$\Rightarrow$ lack *generalization* power

Decision Trees

- Frontier is based on decision rules
- Model has a tree hierarchy
- Each level is a test on attributes



How to create the hierarchy ?

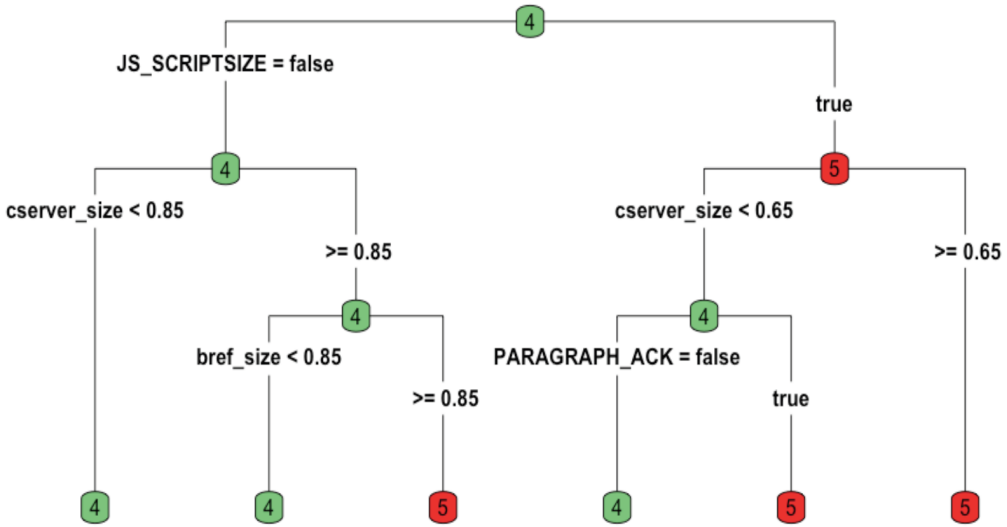
- Multiple heuristics exist
- Most popular : Information gain ; gini

How to create the hierarchy ?

- Multiple heuristics exist
- Most popular : Information gain ; gini

Information gain

- Entropy based heuristic
- What is the attribute that discriminate most ?



Coming back at overfitting

- Few prediction errors on the training set
- Much more on the test set

Coming back at overfitting

- Few prediction errors on the training set
- Much more on the test set

How to avoid it with trees ?

- Constrain the growth of the tree
- Prune the tree

Coming back at overfitting

- Few prediction errors on the training set
- Much more on the test set

How to avoid it with trees ?

- Constrain the growth of the tree
- Prune the tree

Pruning

- based on heuristic
- Impurity measure
- Trade-off between the depth of the tree and number of errors

Exploitation

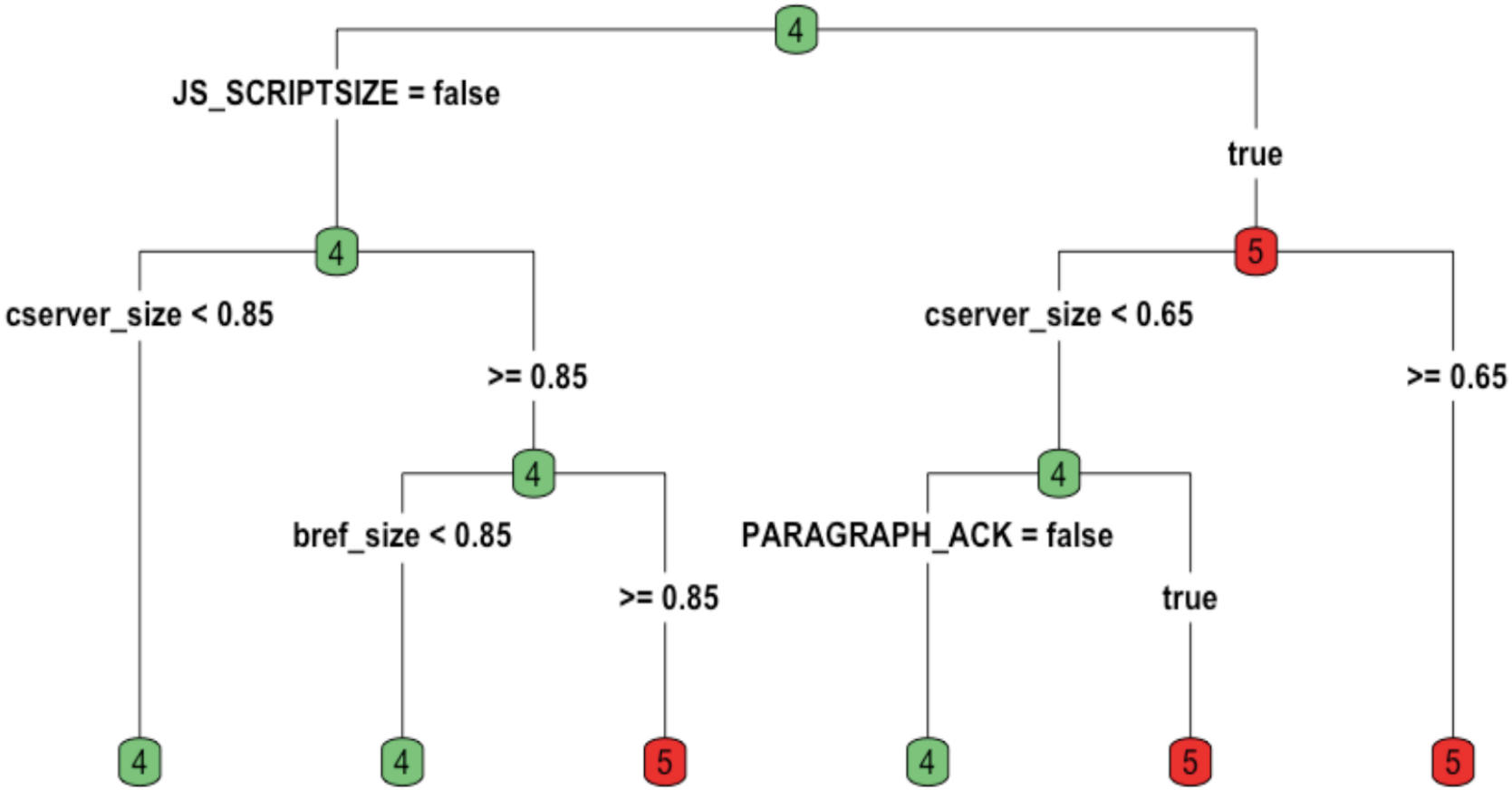
- For each incoming new data
- Begin at the root of the tree
- Follow path depending on results of tests
- The ending leaf gives the class

Exploitation

- For each incoming new data
- Begin at the root of the tree
- Follow path depending on results of tests
- The ending leaf gives the class

Decision rules

- New constraints can be extracted
- Simply follow the path !


4
JS_SCRIPTSIZE = false
true
4
5
cserver_size < 0.85
>= 0.85
cserver_size < 0.65
>= 0.65
4
4
bref_size < 0.85
PARAGRAPH_ACK = false
>= 0.85
true
4
4
5
4
5
5

From Decision to Regression

- Decision = predict a label
- Regression = predict a value

From Decision to Regression

- Decision = predict a label
- Regression = predict a value

Regression Trees

- Similar to Decision Trees
- Problem is more complex

Performance

About # of good predictions and #errors

Performance

About # of good predictions and #errors

Supervised leraning

Confusion Matrix

		Prediction outcome	
		p	**n**
actual value	**p**	True Positive	False Negative
	n	False Positive	True Negative

Performance

About # of good predictions and #errors

Performance

About # of good predictions and #errors

Unsupervised learning

- Mean Squared Error (MSE) and Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE) :
  $\text{MAPE} = \frac{100}{\#observ.} \sum_{i=1}^{\#observ.} \frac{\|expected - predicted\|}{expected}$