Ocean's Big Data Mining, 2014
(Data mining in large sets of complex oceanic data: new challenges and solutions)

8-9 Sep 2014 Brest (France)

Monday, September 8, 2014, 4:00 pm - 5:30 pm

Introduction to data mining. Example of remote sensing image analysis

Prof. Pierre Gançarski, University of Strasbourg – Icube lab.

After a brief introduction about data mining (why-what-how), the talk presents the two type of tasks: prediction tasks which consist in learning a model from data able to predict unknown values, generally class label and description tasks which consist in finding human-interpretable patterns that describe the data. Classification of data, supervised or not, is one of the most used approach in data mining. Then, the talk introduces the most widely used classification and clustering methods.

Applications of these two approaches are mainly illustrated in remote sensing image analysis but a lot of domains are concerned by these methods.

About Pierre Gançarski

Pierre Gançarski is full Professor in Computer Science Department of the ICube laboratory (University of Strasbourg - France).

His research interests concern the complex data mining by hybrid learning multi-classifiers and particularly the unsupervised classification user-centered. It is also interested in studying ways to take into account the domain knowledge in the mining process. His main area (but not the single one) of applications is the classification of remote sensing images.

Data mining and remote sensing images interpretation

Pierre Gançarski

ICube
CNRS - Université de Strasbourg

2014

Contents

1. Data mining

2. Remote sensing image

3. Classification

4. Unsupervised classification

1. Data mining
2. Remote sensing image
3. Classification
4. Unsupervised classification

Data Mining : why ?

Data

- Lots of data is being collected :
  - Web data, e-commerce
  - Purchases at department/grocery stores
  - Bank/Credit Card transactions ; Phone call
  - Skies Telescopes scanning
  - Microarrays to gene expression data
  - Scientific simulations
  - Homics
  - ...

and ... Remote sensing images

Data Mining : why ?

Data

- Lots of data is being collected
- Comprehension/Analysis/Understanding by an human is infeasible for such raw data

Data Mining : why ?

Data

- Lots of data is being collected
- Comprehension/Analysis/Understanding by an human is infeasible for such raw data
- Computers are cheaper and cheaper and more powerful

Data Mining : why ?

Data

- Lots of data is being collected :
- Comprehension/Analysis/Understanding by an human is infeasible for such raw data
- Computers are cheaper and cheaper and more powerful

→ How to extract interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from such data using computer ?

Data Mining : why ?

Data

- Lots of data is being collected :
- Comprehension/Analysis/Understanding by an human is infeasible for such raw data
- Computers are cheaper and cheaper and more powerful

→ How to extract interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from such data using computer ?

Solution

- Data warehousing and on-line analytical processing
- Data mining : extraction of interesting knowledge (rules, regularities, patterns, constraints) from data
- . . .

Data Mining : what ?

Data mining

- Tasks : Knowledge discovery of hidden patterns and insights
- Two kinds of method :
  - Induction-based methods : learn the model, apply it to new data, get the result → Prediction
    - Example : Based on past results, who will pass the DM exam next week and why ?
  - Patterns extraction → Discovery of hidden patterns
    - Example : Based on past results, can we extract groups of students with same behavior ?

Data Mining : how ?

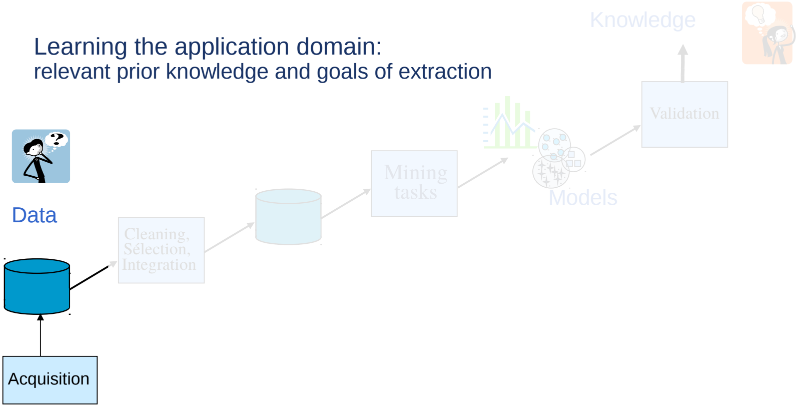
Objective

- Objective of data mining tasks
  - Predictive data mining : Use some variables to learn model able to predict unknown or future values of other variables (e.g., class label) → Induction-based methods
    - Regression, classification, rules extraction
  - Descriptive data mining : Find human-interpretable patterns that describe the data → Patterns extraction
    - Clustering, associations extraction

Data mining : Knowledge Discovery from Data

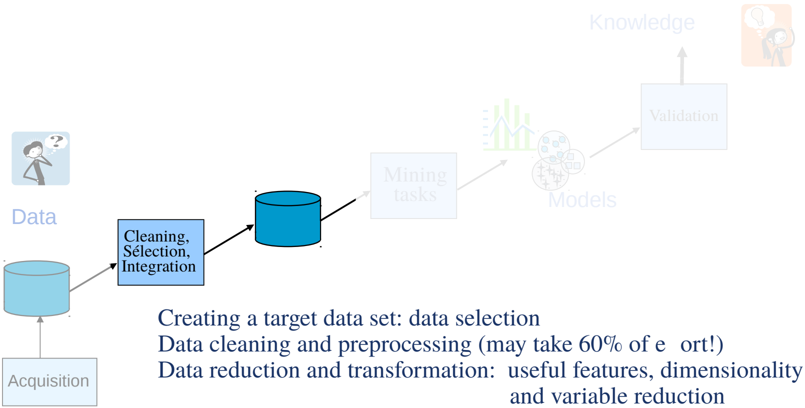
Data Mining : a part of the global KDD process

Learning the application domain:
relevant prior knowledge and goals of extraction



Data mining : Knowledge Discovery from Data

Data Mining : a part of the global KDD process



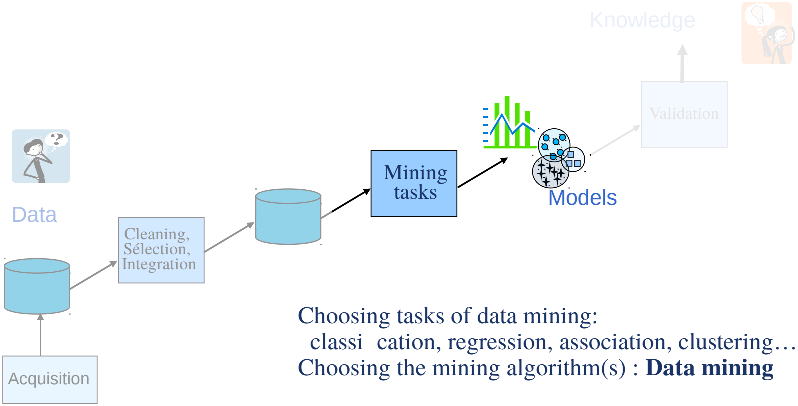
Creating a target data set: data selection

Data cleaning and preprocessing (may take 60% of effort!)

Data reduction and transformation: useful features, dimensionality and variable reduction

Data mining : Knowledge Discovery from Data

Data Mining : a part of the global KDD process



Choosing tasks of data mining:
classification, regression, association, clustering…
Choosing the mining algorithm(s) : **Data mining**

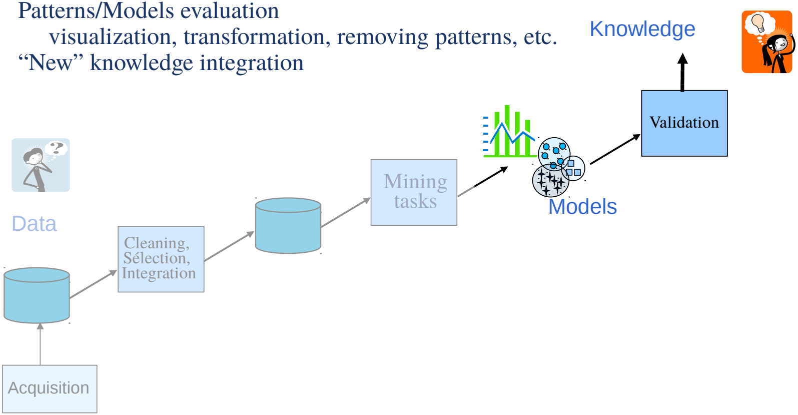
Data mining : Knowledge Discovery from Data

Data Mining : a part of the global KDD process

Patterns/Models evaluation

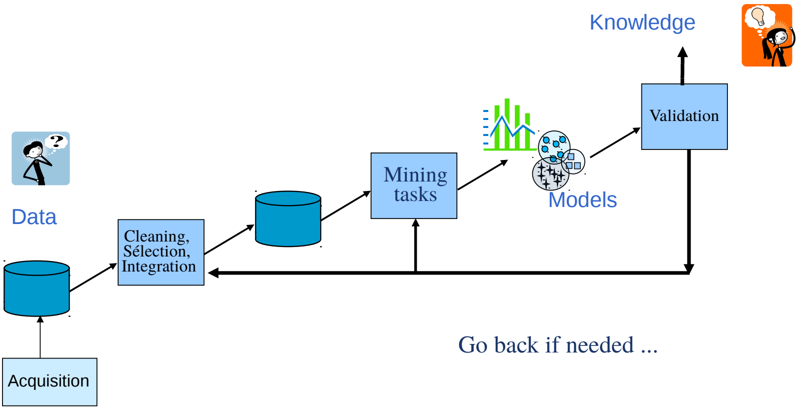
visualization, transformation, removing patterns, etc.

"New" knowledge integration



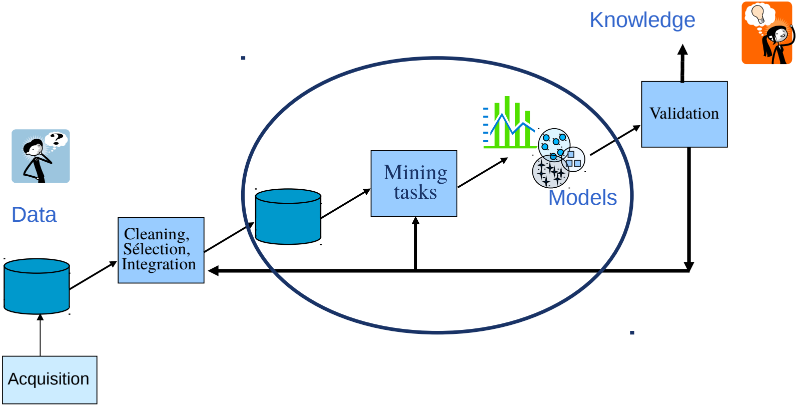
Data mining : Knowledge Discovery from Data

Data Mining : a part of the global KDD process



Data mining : Knowledge Discovery from Data

Data Mining : a part of the global KDD process



Data mining : tasks

Common data mining tasks

- Regression [Predictive]
- Classification [Predictive]
- Clustering [Descriptive]
- Association Rule Discovery [Descriptive]

1. Data mining

2. Remote sensing image

3. Classification

4. Unsupervised classification

Remote sensing image

What's it ?

- Image captured by an aerial or satellite system :
    - optic sensors : spectral responses of various surface covers associated with sunshine
    - radar sensors
    - Lidar
    - ...

Thanks to the LIVE lab (A. Puissant) and the IPGS lab (J.-P. Malet) for providing images.

Remote sensing image

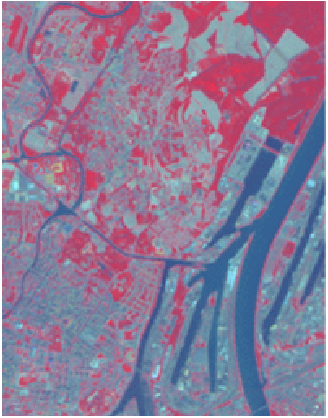
Three dimensions

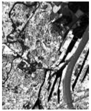
- spatial resolution : surface covered by a pixel (from 300m to few tens of centimetres)
- spectral resolution : number of spectral information (from blue to infrared) corresponding to the number of sensors
- radiometric resolution : linked to the ability to recognize small brightness variations (from 256 to 64000 level)

Remote sensing image

Spatial resolutions

- High spatial resolution (HSR) : 20 or 10m





Green

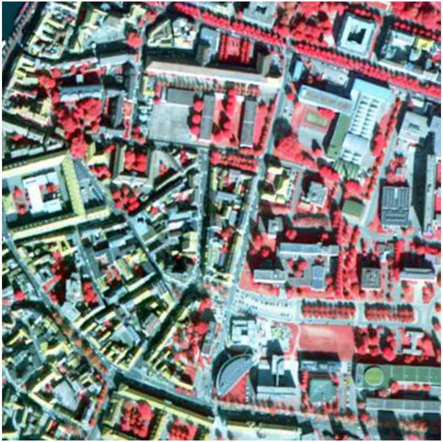
Red

Near infrared (NIR)

Remote sensing image

Spatial resolutions

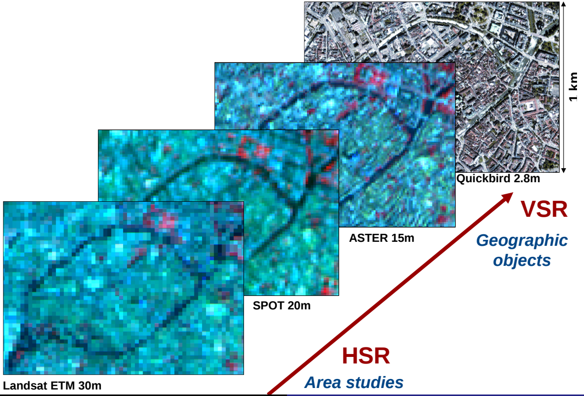
- Very high spatial resolution (VHSR) : from 5m to 0.5m



Remote sensing image

Spatial resolutions

- Level of analysis is linked to the spatial resolution
    - Urban analysis



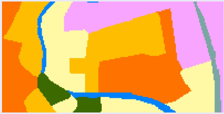
Remote sensing image

Spatial resolutions

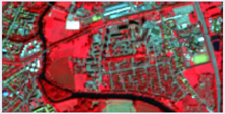
- Level of analysis is linked to the spatial resolution

Examples : Urban blocks

10m : Districts

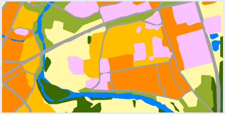


MRS

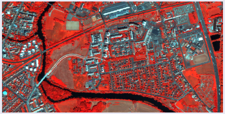


Homogeneous areas

2,4 m : Urban blocks

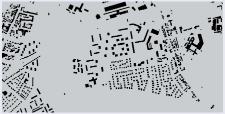


HRS

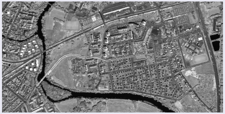


Set of elementary objects spacially organized

60cm : Urban objects



THRS

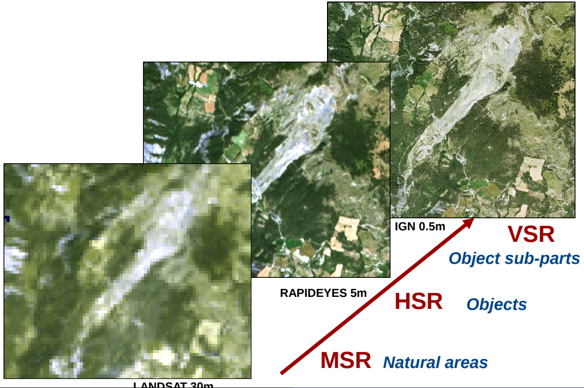


Set of elementary objects with higher heterogeneity

Remote sensing image

Spatial resolutions

- Level of analysis is linked to the spatial resolution
  - Urban analysis
  - Landslide analysis



Remote sensing image

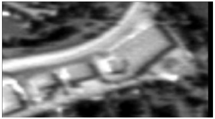
Three dimensions

- spatial resolution : surface covered by a pixel (from 300m to few tens of centimetres)
- spectral resolution : number of spectral information (from blue to infrared) corresponding to the number of sensors
- radiometric resolution : linked to the ability to recognize small brightness variations (from 256 to 64000 level)

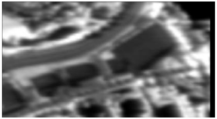
Remote sensing image

Spectral resolutions

- Hight spectral resolution : One hundred of radiometric bands (or more)



band #1



band #22

Aerial view

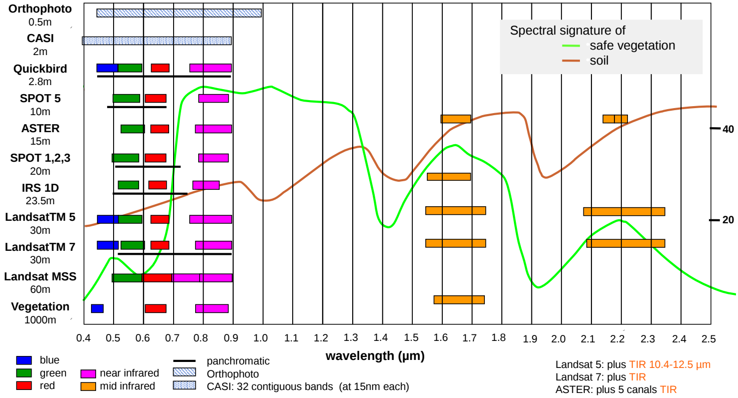
band #29

band #38

Remote sensing image

Image interpretation

- Discrimination between (kinds of) objects can depend on the spectral resolution



Remote sensing image

Three dimensions

- spatial resolution : surface covered by a pixel (from 300m to few tens of centimetres)
- spectral resolution : number of spectral information (from blue to infrared) corresponding to the number of sensors
- radiometric resolution : linked to the ability to recognize small brightness variations (from 256 to 64000 level)

Image interpretation

Semantic gap

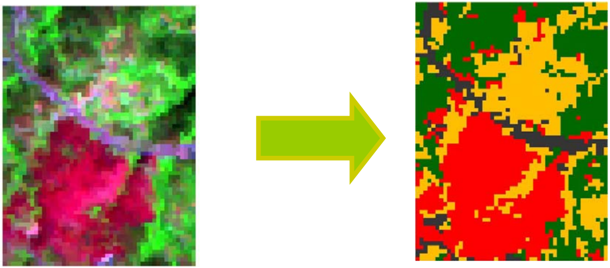
- There are differences between the *visual* interpretation of the spectral information and the semantic interpretation of the pixels
- The semantic is not always explicitly contained in the image and depends on domain knowledge and on the context.

⟹ This problem is known as the *semantic gap* and is defined as the lack of concordance between low-level information (*i.e.* automatically extracted from the images) and high-level information (*i.e.* analyzed by geographers)

Pixels or regions

Per-pixel analysis

- Pixels are analyzed according only their radiometric responses (with possibly some indexes of texture and/or immediate neighbourhood)



Pixels or regions

Per-pixel analysis

- Pixels are classified analyzed only their radiometric responses (with possibly some indexes of texture and/or immediate neighbourhood)

Efficient with low resolutions but ...

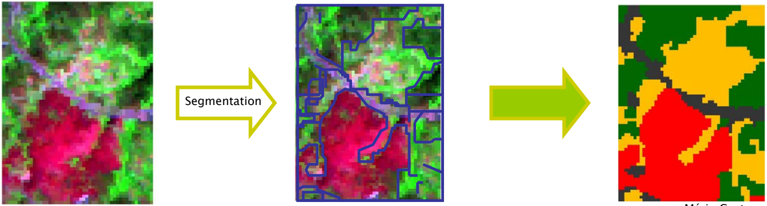
- In (V)HSR images, a pixel is only almost any cases, a small sub-part of the thematic object
- The object to be analyzed are composed of a lot of pixels (from few tens to few hundred or more).

→ New approach : Object-based Image Analysis (OBIA)

Pixels or regions

Object-based Image Analysis

1. Segmentation of the image : grouping neighbouring pixels according a given homogeneous criterion
2. Characterization of these segments with supplemental radiometric, textural, spatial features, . . .
   $\rightarrow$ object (or region) = segment + set of features
3. Analysis of these regions



Mário Caetano

Pixels or regions

Region-based classification

1. Segmentation of the image : grouping neighbouring pixels according a given homogeneous criterion
2. Characterization of these segments with supplemental radiometric, textural, spatial features, . . .
   $\rightarrow$ region = segment + set of features
3. Classification of these regions

Two main problems

- Segmentation
- Feature identification and selection

Pixels or regions

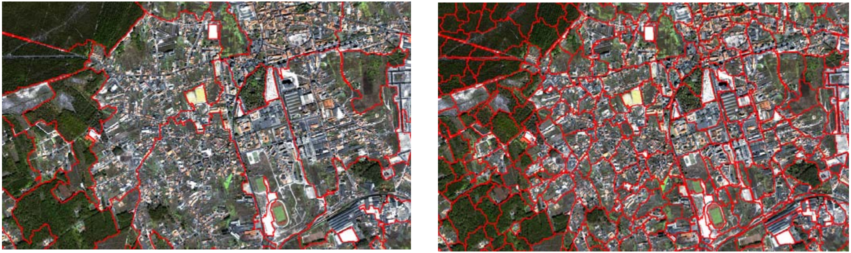
Image segmentation

- Segmentation is the division of an image into spatially continuous, disjoint and homogeneous areas, i.e. the segments.
- Segmentation of an image into a given number of regions is a problem with a large number of possible solutions.
- There are no “right”, “wrong” or “better” solutions but instead “meaningful” and “useful” heuristic approximations of partitions of space.

Pixels or regions

Image segmentation

- These two segmentation is intrinsically right.
  The choice depends on their usefulness in the next step (e.g., classification)



Mário Caetano

Pixels or regions

Feature identification and selection

- What type of features can be used for geographic object classification ? There a an infinity of such features.
- How to select the best features for class discrimination ?

→ How to select the smaller number of features without sacrificing accuracy ?

Image interpretation

Solutions

- To bridge this semantic gap, a lot of approaches can be used :
  - Human visual interpretation : untractable with the new kinds of images (VHSR especially)
  - Per pixel classification (supervised or not) : Automatic or semi-automatic labelisation of the pixels according to thematic classes
  - Segmentation and region classification : Construction of sub-parts of the image and labelisation of these (possibly by a classification process)
  - ...

1. Data mining
2. Remote sensing image
3. Classification
4. Unsupervised classification

Principles of classification

Induction

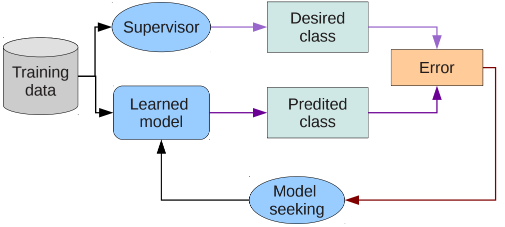
- Classification is based on induction principle
  - Deductive reasoning is truth-preserving :
    - All human have a heart
    - All professors are human (for the moment)
    - Therefore, all professors have a heart
  - Induction reasoning adds information
    - All professor observed so far have a heart.
    - Therefore, all professors have a heart.

Principles of classification

Supervised classification

- Supervised classification is the task of inferring a model from labelled training data : each example is a pair consisting of an input object and a desired class.

⇒ The classes to be learned are known a priori

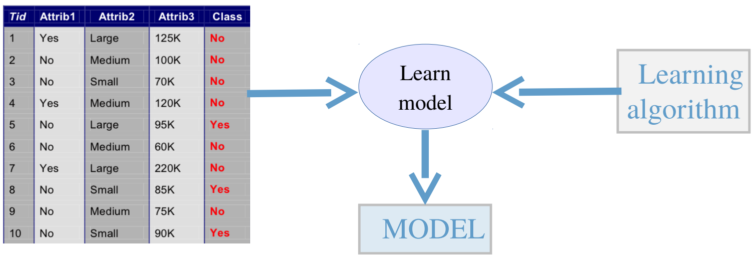


- Classical methods : Support vector machine, Decision tree, Artificial neural network. . .

Classification schema

Offline : Induction (training)

- Define a `Training set` : Each record contains a set of attributes, one of the attributes is the class (class label).
- Find a model for class attribute as a function of the values of other attributes.



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Learn model

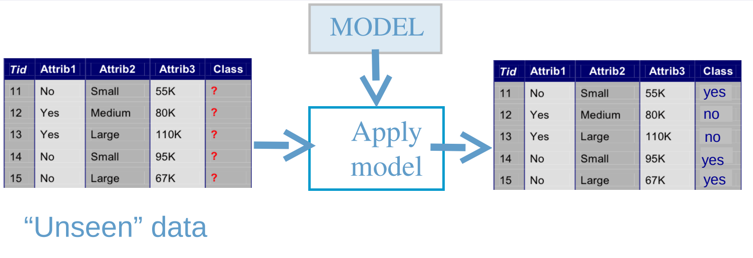
Learning algorithm

MODEL

Classification schema

Inline : Deduction (Generalization or Prediction)

- Using the model, give a class label to unseen records



Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	yes
12	Yes	Medium	80K	no
13	Yes	Large	110K	no
14	No	Small	95K	yes
15	No	Large	67K	yes

"Unseen" data

Classification evaluation

How to evaluate the performance of a model ?

- Use of a confusion matrix :
  - Example : N data and two classes + and -

		Predicted class	
		+	-
Actual class	+	A	B
	-	C	D

  - A : True positive (TP) ; B : False negative (FN) ; C : False positive (FP) ; D : True negative (TN)
  - $N = A + B + C + D$

Classification evaluation

Accuracy

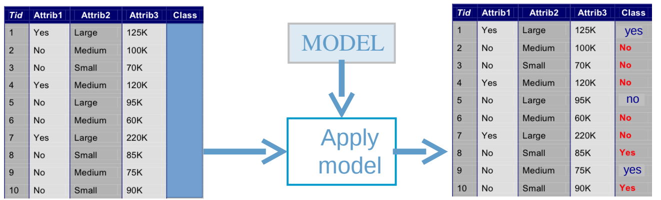
- Defined as the ratio of well-classified objects :

$$\frac{A+D}{A+B+C+D} = \frac{TP+TN}{N}$$

Classification evaluation

Accuracy

- Example : We can apply the learned model to the data (without using class labels !)



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	
2	No	Medium	100K	
3	No	Small	70K	
4	Yes	Medium	120K	
5	No	Large	95K	
6	No	Medium	60K	
7	Yes	Large	220K	
8	No	Small	85K	
9	No	Medium	75K	
10	No	Small	90K	

MODEL

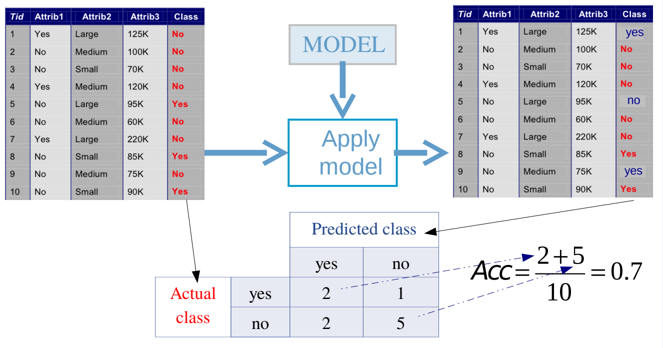
Apply model

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	yes
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	no
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	yes
10	No	Small	90K	Yes

Classification evaluation

Accuracy

- Example : We can apply the learned model to the data (without using class labels !)



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	yes
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	no
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	yes
10	No	Small	90K	Yes

		Predicted class	
		yes	no
Actual class	yes	2	1
	no	2	5

$$Acc = \frac{2+5}{10} = 0.7$$

- Conclusion : 30% of training error

Classification evaluation

Accuracy

- Example : We can apply the learned model to the data (without using class labels !)
- Conclusion : 30% of training error
- Suppose we redo training (with new parameters for instance) until this error is zero or near.

⇝ Question : Is the model right now ?

Classification evaluation

Accuracy

- Example : We can apply the learned model to the data (without using class labels !)
- Conclusion : 30% of training error
- Suppose we redo training (with new parameters for instance) until this error is zero or near.

⇝ Question : Is the model right now ?

- Not necessary : the most important is that the model correctly classifies unknown data

Classification evaluation

Accuracy

- Example : We can apply the learned model to the data (without using class labels !)
- Conclusion : 30% of training error
- Suppose we redo training (with new parameters for instance) until this error is zero or near.

⇝ Question : Is the model right now ?

- Not necessary : the most important is that the model correctly classifies unknown data

⇝ Question : How to calculate accuracy in this case without any information about actual classes ?

Classification evaluation

Accuracy

- How to calculate accuracy without any information about actual classes ?
  1. Ask the expert
  2. Use the labeled data in a different way by keeping some of them to test the model

Classification evaluation

Cross-validation

- Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it (often 2/3 - 1/3).
- N-fold cross-validation :
  1. The given data set is divided into N subset
  2. The model is learn using (N-1) subsets
  3. The remaining subset is used to validation
  4. The operation (step 2) is repeated N times (with a different test subset each time).
  5. The best model is kept.

Classification evaluation

Accuracy limitation

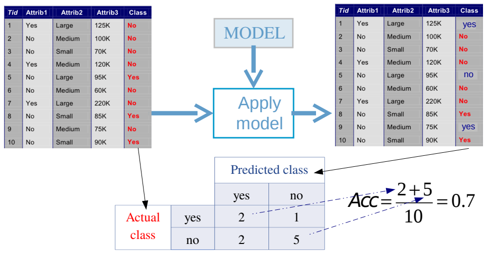
- The accuracy is very sensible to the skew of data
- Number of examples - = 9990 ; Number of examples $+$ = 10
  If model predicts everything to be class -, accuracy is $9990/10000 = 99.9$ %

Precision/Recall

- $P(c) = \frac{\#(c|c)}{\#(c|c)+\#(c|\bar{c})}$ $\rightsquigarrow$ ratio of objects classified as c which actually belong to class c
- $R(c) = \frac{\#(c|c)}{\#(c|c)+\#(\bar{c}|c)}$ $\rightsquigarrow$ ratio of objects belonging to actual class c which are classified as c

Classification evaluation

Precision/Recall



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	yes
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	no
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	yes
10	No	Small	90K	Yes

		Predicted class	
		yes	no
Actual class	yes	2	1
	no	2	5

$$Acc = \frac{2+5}{10} = 0.7$$

- $P(yes) = \dfrac{\#(yes|yes)}{\#(yes|yes)+\#(yes|no)} = \dfrac{2}{2+2} = 1/2$

  $R(yes) = \dfrac{\#(yes|yes)}{\#(yes|yes)+\#(no|yes)} = \dfrac{2}{2+1} = 2/3$
- $P(no) = \dfrac{\#(no|no)}{\#(no|no)+\#(no|yes)} = \dfrac{5}{5+1} = 5/6$

  $R(no) = \dfrac{\#(no|no)}{\#(no|no)+\#(yes|no)} = \dfrac{5}{5+2} = 5/7$

Classification

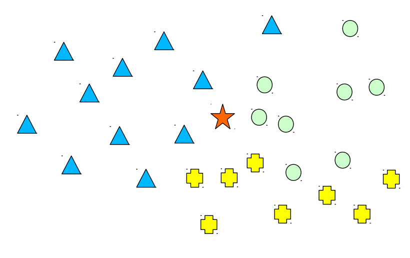
A lot of classification techniques

- The most popular :
  - K-Nearest-Neighbor
  - Decision Tree based Methods
  - Bayesian classifier : not directly usable in RS analysis
  - Rule-based Methods : not directly usable in RS analysis
  - Artificial Neural Networks : to long to explain (sorry)
  - Support Vector Machines → tomorrow

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors

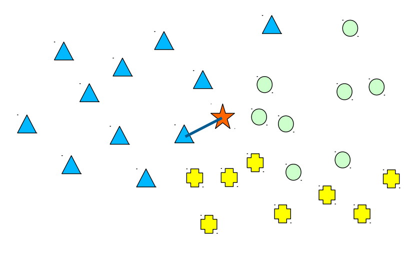


- What is the class of red star ?

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors

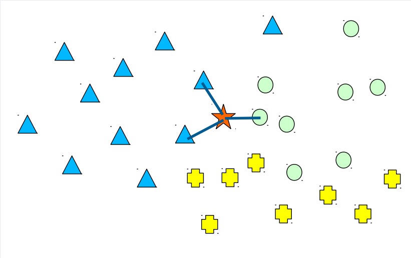


- k=1 : blue triangle

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors

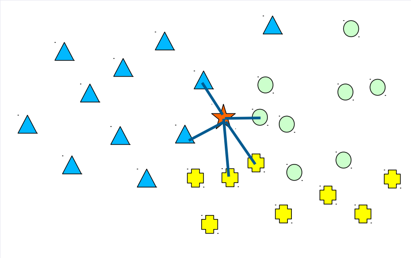


- k=1 : blue triangle ; k=3 : blue triangle ;

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors

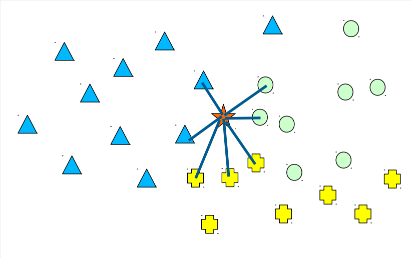


- k=1, k=3 : blue triangle ; k=5 : undefined

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors

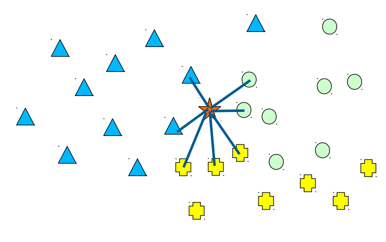


- k=1, k=3 : blue triangle ; k=5 : undefined ; k=7 : yellow cross

K-Nearest-Neighbor

Principle

- Use of a similarity measure (or distance) between data
- The label of unseen data is set according to label of its K nearest neighbors



- k=1, k=3 : blue triangle ; k=5 : undefined ; k=7 : yellow cross
- k = N ?

K-Nearest-Neighbor

Characteristics

- Lazy learner : It does not build models explicitly
- Classifier new data is relatively expensive
- Too high influence (unstability) of the K parameter
- Need of a similarity measure (or distance) between data

Simarity measure

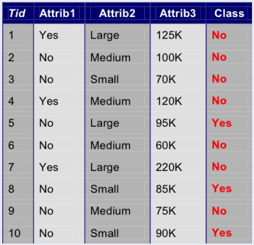
- Per-pixel classification : Euclidean distance between radiometric values $\sqrt{(XS1_i - XS1_j)^2 + \cdots + (XS3_i - XS3_j)^2}$
- Object-oriented approaches : depends on types of features $\rightsquigarrow$ Can be difficult to define

Decison Trees

Principle

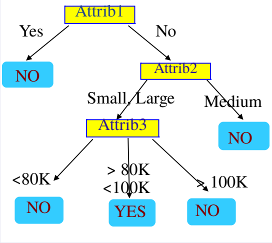
- Decision trees are rule-based classifiers that consist of a hierarchy of decision points (the nodes of the tree).

Training



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

≡ ▷



Decison Trees

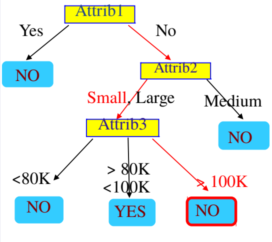
Principle

- Decision trees are rule-based classifiers that consist of a hierarchy of decision points (the nodes of the tree).

Prediction

Attr1 = No
Attrib2 = Small
Attrib3 = 130K
Class = ? ?

≡ ▷



Decison Trees

Training

- How to build a such tree from the training data ? → Recursive partitioning
- At the beginning, all the records are in the root node (considered as the current node C).
- General procedure :
  - If C only contains records from the same class c_t then C become the leaf labeled c_t
  - If C contains records that belong to more than one class, use an attribute test to split the data into smaller subsets (associated each to a children node).
  - Recursively apply the procedure to each children node

Decison Trees

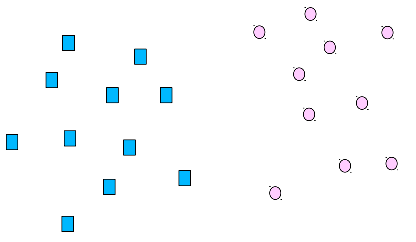
Training

- How to specify the attribute test condition ?
  - Depends on the attribute types : nominal,ordinal, continuous
  - Depends on number of ways to split : binary / multiple
- How to determine the best split ?
  - "Heterogenity" indexes :
    Gini Index $G(C) = 1 - \Sigma[p(c_i|C)]^2$;
    Entropy : $E(C) = -\Sigma p(c_i|C) log_2 p(c_i|C)$
  - Misclassification error
  - ...

Classification

Hyperplan-based methods

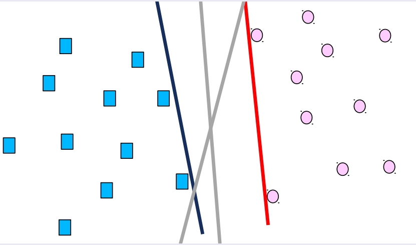
- Principle : Given a to classes training dataset , find a hyperplan which separates data into the two classes



Classification

Hyperplan-based methods

- Principle : Given a to classes training dataset , find a hyperplan which separates data into the two classes



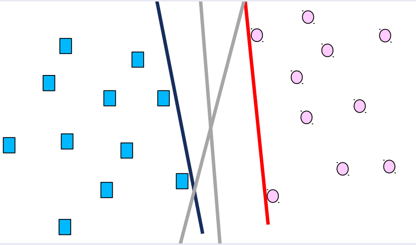
- Where is really the boundary ?

⇝ Overfitting appear when the model would exactly fit to the training data

Classification

Hyperplan-based methods

- Principle : Given a to classes training dataset , find a hyperplan which separates data into the two classes

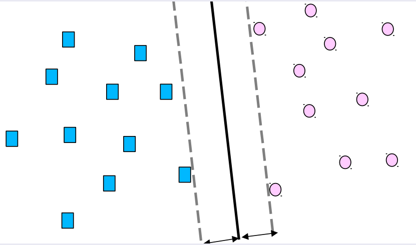


- Where is really the boundary ?
  - The red hyperplan probably overfits the “Pink circle” class while the blue one the “Blue square” one

Classification

Hyperplan-based methods

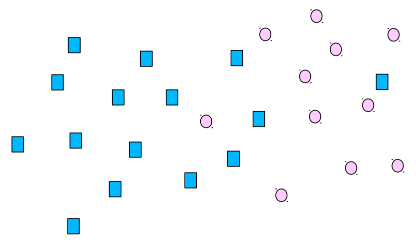
- Methods :
  - Artificial neural network find one of such hyperplan
  - Support Vector Machine find “the best one”



Classification

Hyperplan-based methods

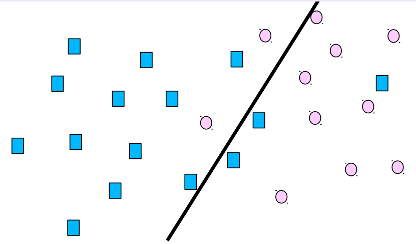
- What happen if the dataset is not linearly separable ?



Classification

Hyperplan-based methods

- What happen if the dataset is not linearly separable ?

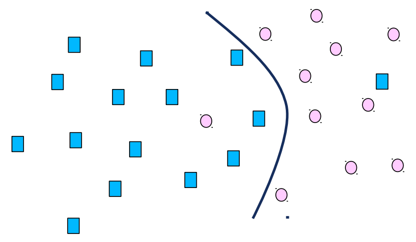


- The model is too simple ⇒ underfitting

Classification

Hyperplan-based methods

- What happen if the dataset is not linearly separable ?

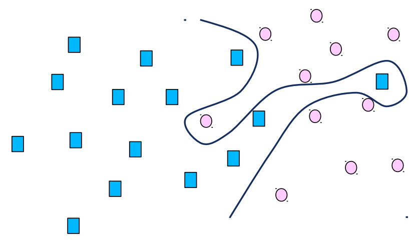


- Complexify the model ?

Classification

Hyperplan-based methods

- What happen if the dataset is not linearly separable ?

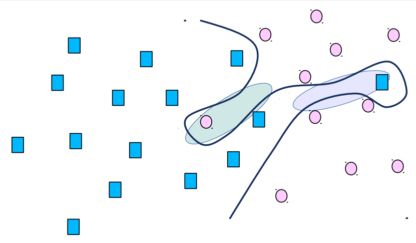


- Complexify the model : more and more ?

Classification

Hyperplan-based methods

- What happen if the dataset is not linearly separable ?

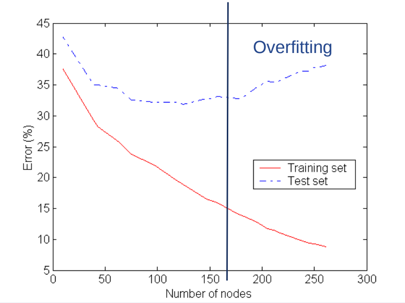


- Problem : if the blue square is a noise (or an outlet) all the points in light blue area will be classified as "blue square" → Test error can increase → Overfitting

Classification

Hyperplan-based methods

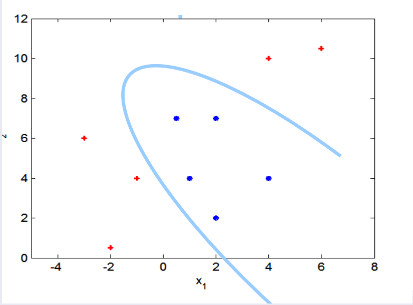
- Experiments show that probability of overfitting increases with complexity.
- For instance, complexity of a decision tree increases with the number of node ⇝ When stop the learning ?



Classification

Hyperplan-based methods

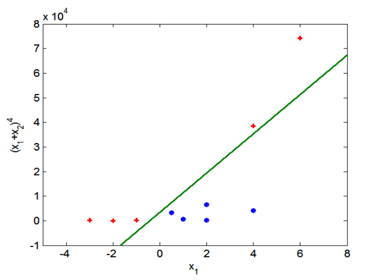
- What happen if the boundary exists but is no linear ?



Classification

Hyperplan-based methods

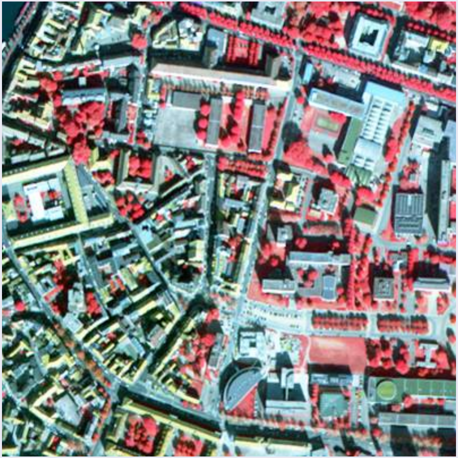
- What happen if the boundary exists but is no linear ?
- Transform data by projecting them into higher dimensional space in which they are linearly separable



⇝ Kernel-based methods : Support Vector Machine

Classification and images

Supervised classification : case of VHSR



What kind of object as training data ?

- pixels ?

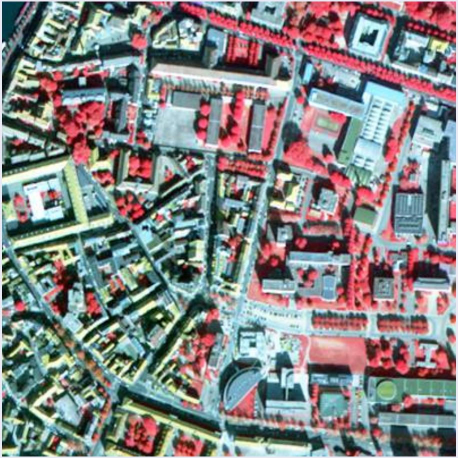
→ Represent only small parts of real objects

- geographic object ?

→ Need to construct candidate segments before learning.

Classification and images

Supervised classification : case of VHSR



How many classes ?

- 10 ?
  - Road
  - Building . . .
- 50 ?
  - Road
  - Street
  - Red car
  - Blue car
  - Lighted (half) roof
  - Shadowed roof . . .

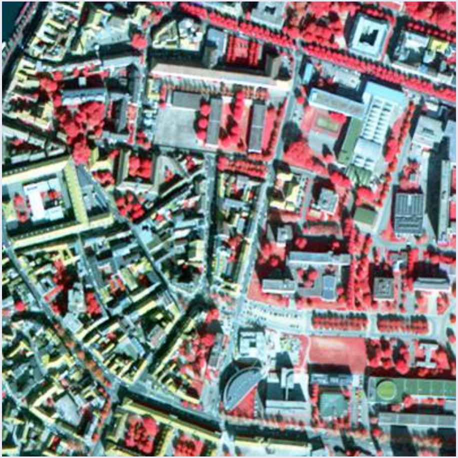
Classification and images

Supervised classification : case of VHSR

How to give enough examples by class with high number of classes ?

Classification and images

Supervised classification : case of VHSR



Supervised approaches
are mainly used
to pixels classification
in low resolution images

1 Data mining

2 Remote sensing image

3 Classification

4 Unsupervised classification

Unsupervised classification

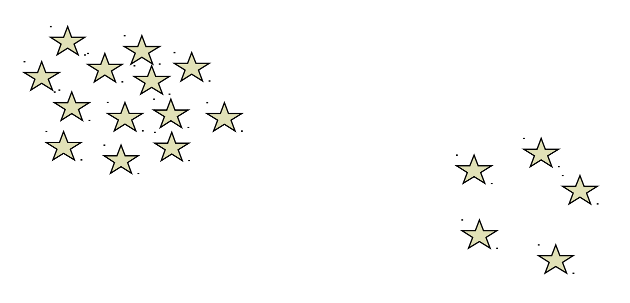
Supervised vs. Unsupervised learning

- Supervised learning : discover patterns in the data that relate data attributes with a target (class) attribute.
  - These patterns are then utilized to predict the values of the target attribute in future data instances.
- Unsupervised learning : The data have no target attribute
  - Explore the data to find some intrinsic structures or hidden properties.

Unsupervised classification

Clustering

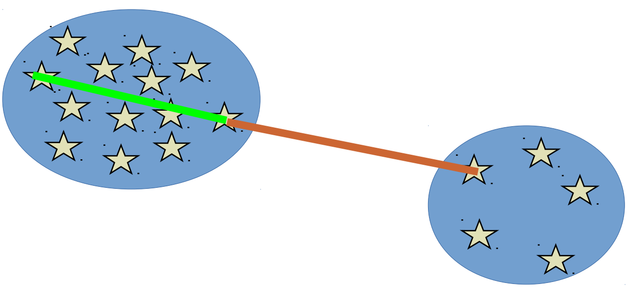
- Clustering is a technique for finding similarity groups in data, called clusters :
    -
    -



Unsupervised classification

Clustering

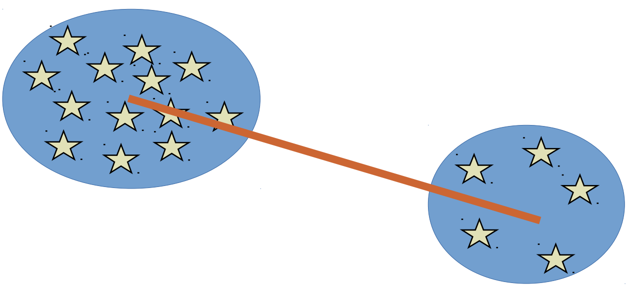
- Clustering is a technique for finding similarity groups in data, called clusters :
    - Homogeneous groups such that two objects of the same class are more similar two objects of different classes
    -



Unsupervised classification

Clustering

- Clustering is a technique for finding similarity groups in data, called clusters :
  - Homogeneous groups such that two objects of the same class are more similar two objects of different classes
  - Homogeneous groups such as dissimilarity/distances between groups are highest



Unsupervised classification

Unsupervised classification : clustering

- Clustering is often called an unsupervised learning task as no class values denoting an a priori grouping of the data instances are given, which is the case in supervised learning.
- Due to historical reasons, clustering is often considered synonymous with unsupervised learning.
    - Association rule mining is also unsupervised
- Typical applications
    - As a stand-alone tool to get insight into data distribution
    - As a preprocessing step for other algorithms
- Each cluster would be assigned to a thematic class by the expert.

Unsupervised classification

Unsupervised classification : clustering

- Partitioning : Construct various partitions and then evaluate them by some criterion
- Hierarchical : Create a hierarchical decomposition of the set of objects using some criterion
- Model-based : Hypothesize a model for each cluster and find best fit of models to data
- Density-based : Guided by connectivity and density functions

Unsupervised classification

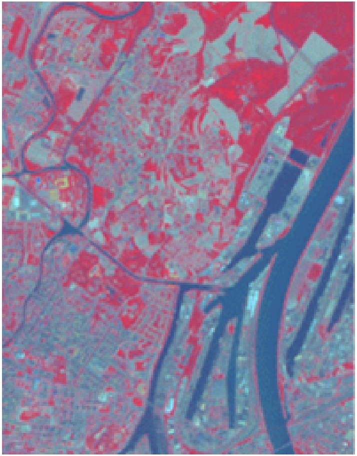
A well-known partitioning algorithm : Kmeans

- Objective : Minimization of intraclass inertia
  ⇝ Given K, find a partition of K clusters that optimizes the chosen partitioning criterion
- Process :
  1. Choice of the number K of clusters
  2. Random choice of K centroids (seed) in the data space
  3. Iteration :
     1. Assign each pixel to the cluster that has the closest centroid
     2. Recalculate the positions of the K centroids
  4. Repeat Step 3 until the centroids no longer move

Unsupervised classification

A well-known clustering algorithm : Kmeans

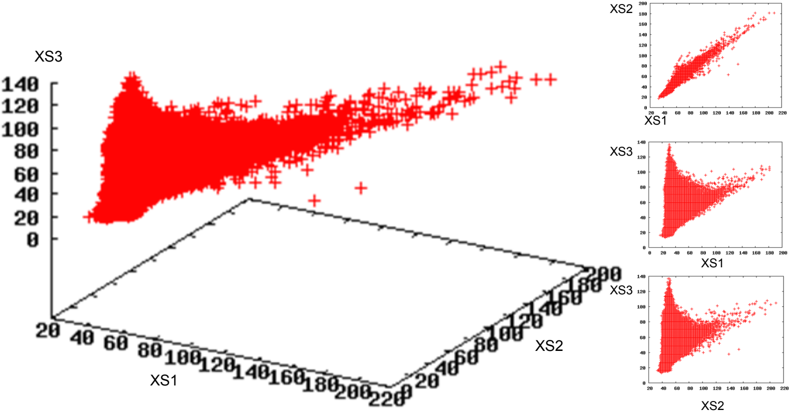
Example on SPOT image



Unsupervised classification

A well-known clustering algorithm : Kmeans

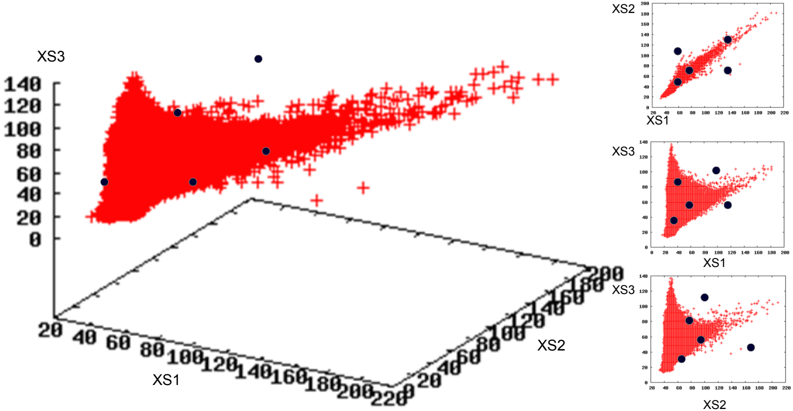
The data space corresponding to the Spot image



Unsupervised classification

A well-known clustering algorithm : Kmeans

Random choice of K = 5 centroids (seed) in the data space



Unsupervised classification

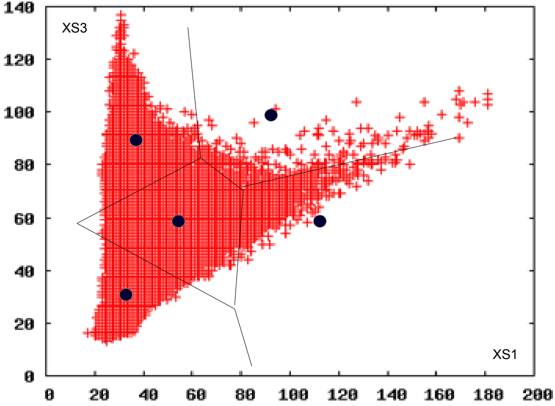
A well-known clustering algorithm : Kmeans

- Objective : Minimization of intraclass inertia
- Process :
  1. Choice of the number K of clusters
  2. Random choice of K centroids (seed) in the data space
  3. Iteration :
     1. Assign each pixel to the cluster that has the closest centroid
     2. Recalculate the positions of the K centroids
  4. Repeat Step 3 until the centroids no longer move

Unsupervised classification

A well-known clustering algorithm : Kmeans

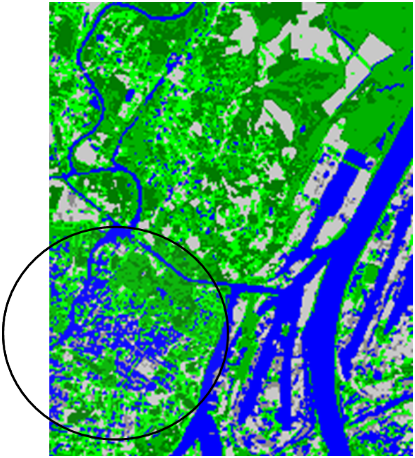
Assign each pixel to the cluster that has the closest centroid



Unsupervised classification

A well-known clustering algorithm : : Kmeans

Each pixel is colorized according to the clusters is belong to.



Unsupervised classification

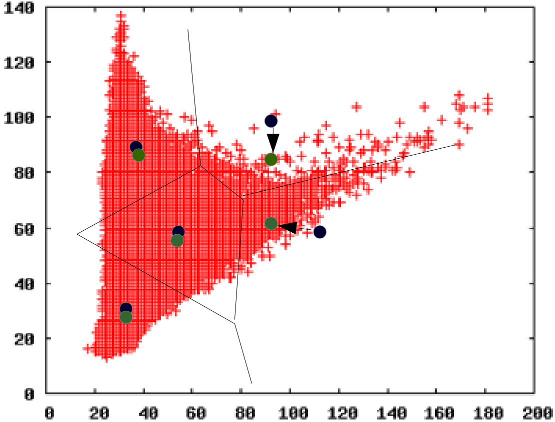
A well-known clustering algorithm : Kmeans

- Objective : Minimization of intraclass inertia
- Process :
  1. Choice of the number K of clusters
  2. Random choice of K centroids (seed) in the data space
  3. Iteration :
     1. Assign each pixel to the cluster that has the closest centroid
     2. Recalculate the positions of the K centroids
  4. Repeat Step 3 until the centroids no longer move

Unsupervised classification

A well-known clustering algorithm : Kmeans

Recalculate the positions of the K centroids



Unsupervised classification

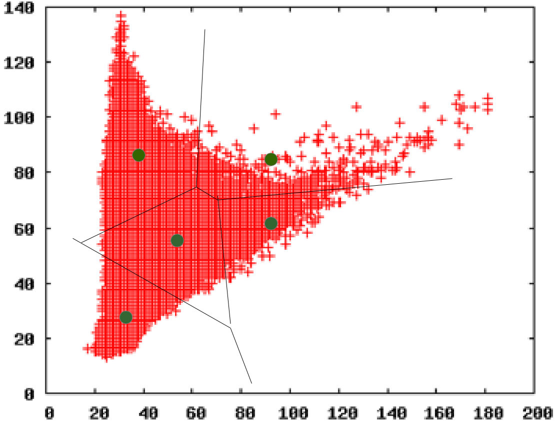
A well-known clustering algorithm : Kmeans

- Objective : Minimization of intraclass inertia
- Process :
  1. Choice of the number K of clusters
  2. Random choice of K centroids (seed) in the data space
  3. Iteration :
     1. Assign each pixel to the cluster that has the closest centroid
     2. Recalculate the positions of the K centroids
  4. Repeat Step 3 until the centroids no longer move

Unsupervised classification

A well-known clustering algorithm : Kmeans

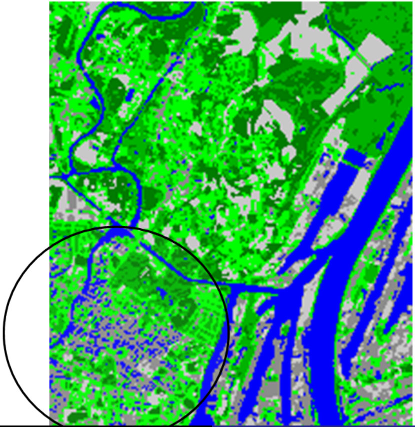
Assign each pixel to the cluster that has the closest centroid



Unsupervised classification

A well-known clustering algorithm : Kmeans

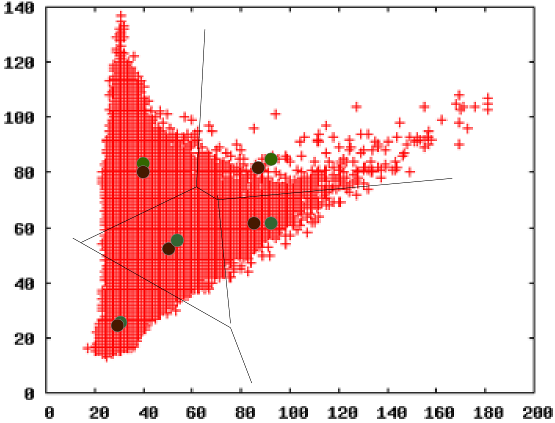
On SPOT image, each pixel is colorized according to the clusters is belong to.



Unsupervised classification

A well-known clustering algorithm : Kmeans

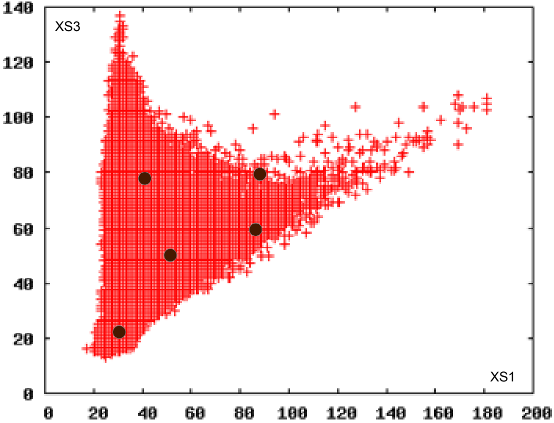
Recalculate the positions of the K centroids



Unsupervised classification

A well-known clustering algorithm : Kmeans

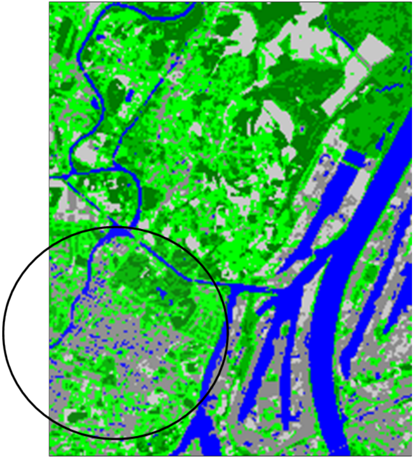
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

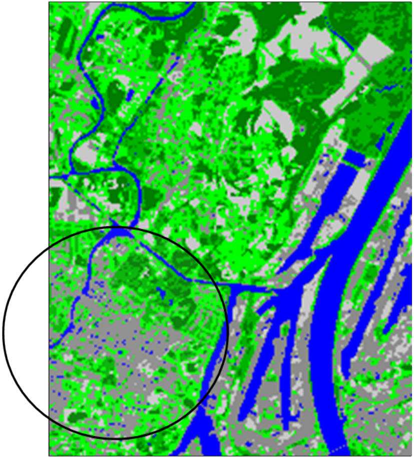
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

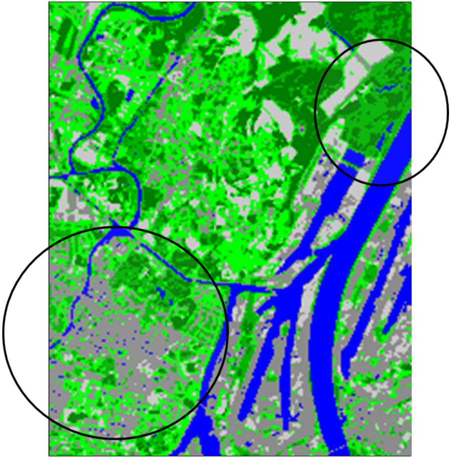
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

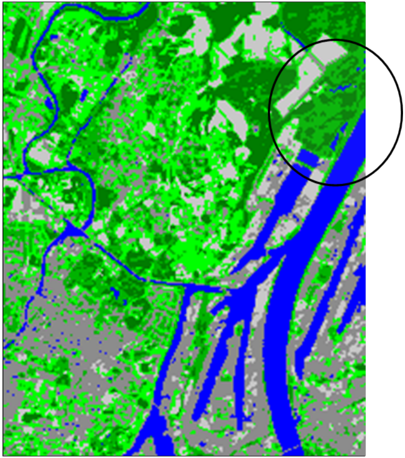
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

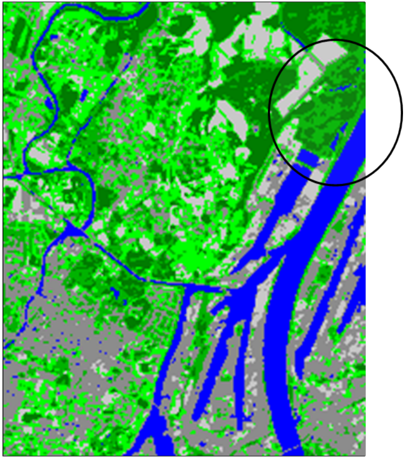
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

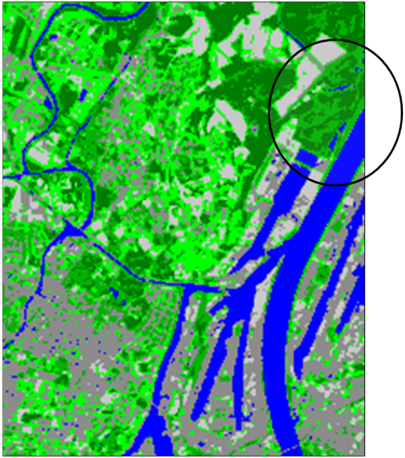
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

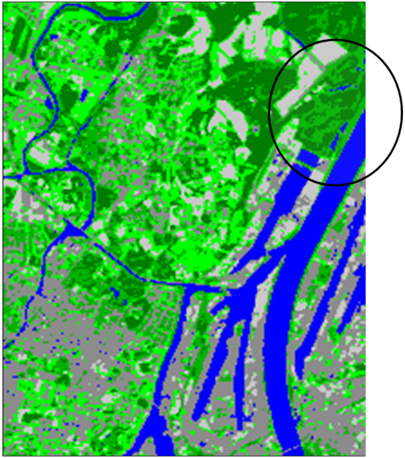
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

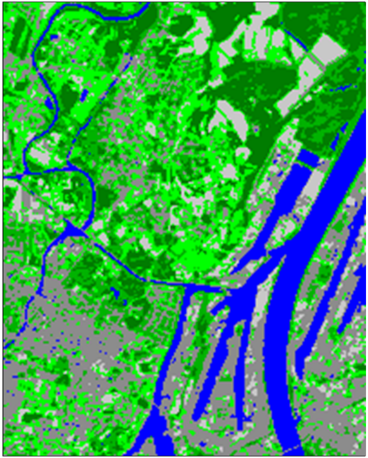
and so on ...



Unsupervised classification

A well-known clustering algorithm : Kmeans

until the centroids do not move



Unsupervised classification

Kmeans : Weaknesses

- The algorithm is only applicable iff mean is defined.
- The user needs to specify K.
- Kmeans is sensitive to outliers
- Kmeans is strong sensitive to initialization
- Kmeans is not suitable to discover clusters with non-convex shapes

$\rightsquigarrow$ Despite weaknesses, Kmeans is still one of the most popular algorithms due to its simplicity and efficiency

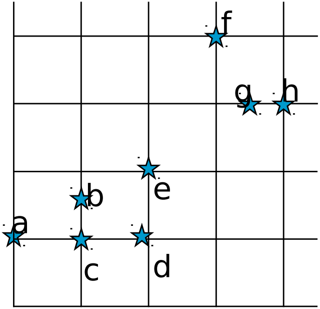
Unsupervised classification

Hierarchical clustering

- Agglomerative (bottom up) clustering :
  - merges the most similar (or nearest) pair of clusters or objects
  - stops when all the data objects are merged into a single cluster
- Divisive (top down) clustering : It starts with all data points in one cluster, the root.
  - Splits the root into a set of child clusters. Each child cluster is recursively divided further
  - Stops when only singleton clusters of individual data points remain, i.e., each cluster with only a single point

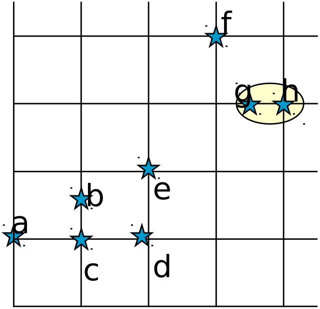
Unsupervised classification

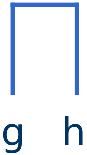
Ascendant hierarchical clustering



Unsupervised classification

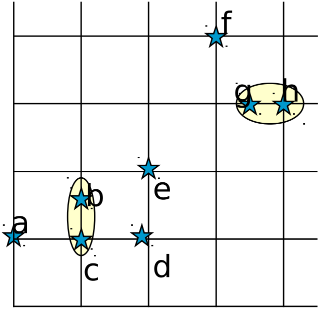
Ascendant hierarchical clustering

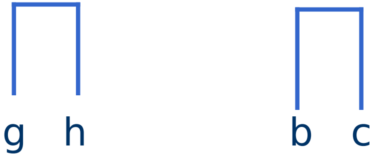




Unsupervised classification

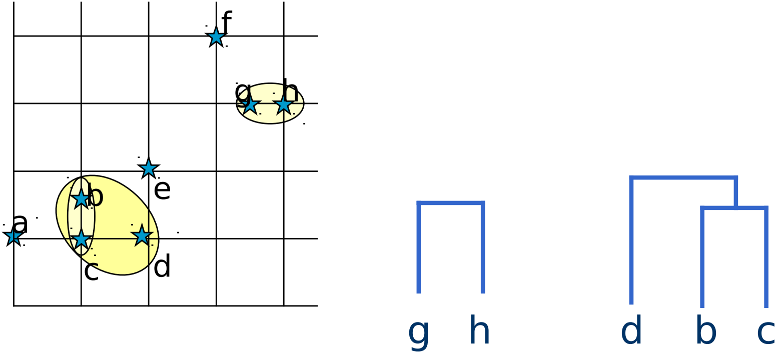
Ascendant hierarchical clustering





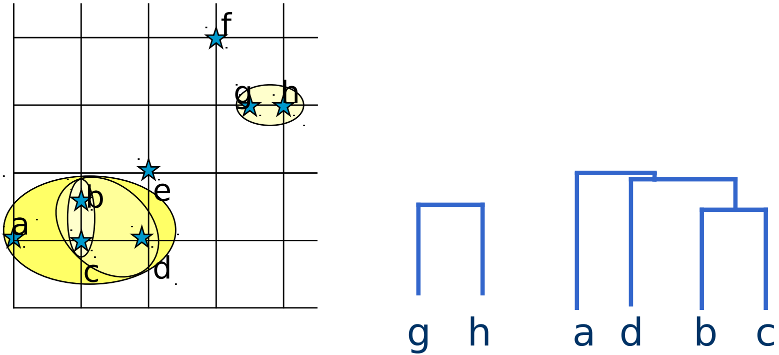
Unsupervised classification

Ascendant hierarchical clustering



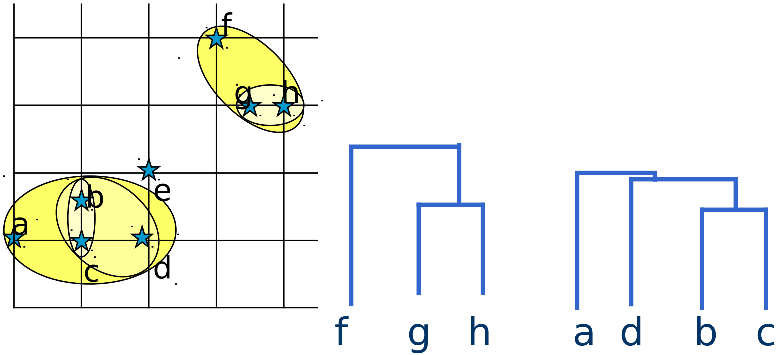
Unsupervised classification

Ascendant hierarchical clustering



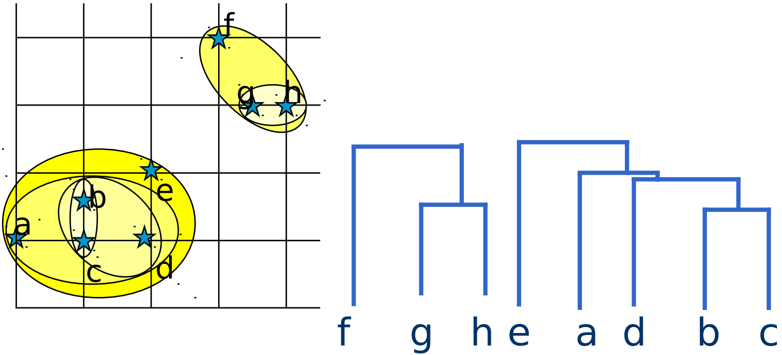
Unsupervised classification

Ascendant hierarchical clustering



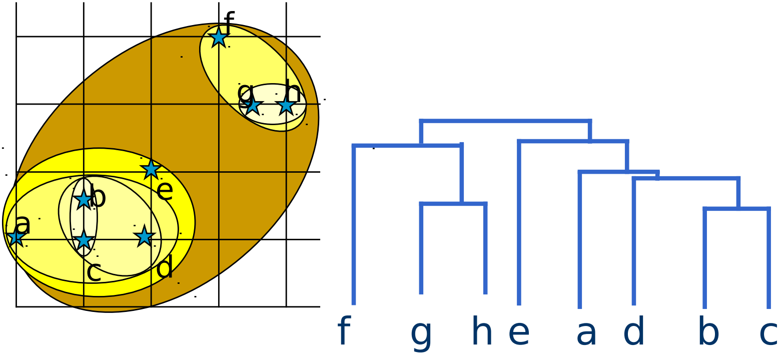
Unsupervised classification

Ascendant hierarchical clustering



Unsupervised classification

Ascendant hierarchical clustering



Unsupervised classification

Ascendant hierarchical clustering

- Avantage : No need of average method
- Weakness : Algorithm cost (two-by-two distance evaluate)
- Often preceded by partitioning algorithm to reduce the dataset size