CIS501 – Lecture 3

Woon Wei Lee
Fall 2013, 10:00am-11:15am,
Sundays and Wednesdays

Masdar INSTITUTE

For today:

- Administrative stuff
- Core Data Mining concepts
  - Coping with the “Curse of Dimensionality”
  - Supervised vs Unsupervised

Coping with the curse..

- **Basic Intuition**:
  - Data could be high (~infinite) dimensional, but the number of relevant *generating processes* is rarely that high.
  - Notion of “intrinsic dimensionality” (which we'd discussed last week → ***Complexity***
- **Broad strategy – eliminate complexity!**
  - Remove unnecessary dimensions
  - Consolidate along informative directions
  - Constrain degrees of freedom
- **Two approaches:**
  - Work with the model...
  - .. or with the data

Coping.. (fix the model)

- **Approach 1**: “Regularization”
  - Aims to alleviate ill-conditioning by introducing constraints to the model
  - Aim: reduce the degrees of flexibility of the resulting models, and (hopefully) eliminate less likely solutions
  - Two common ways of doing this:
    - Constraints on the types of models which are permissible
    - Penalize parameters which correspond to unnecessarily complex models.
  - Can be viewed as a means of incorporating *prior knowledge* into a model.

Fix the data..

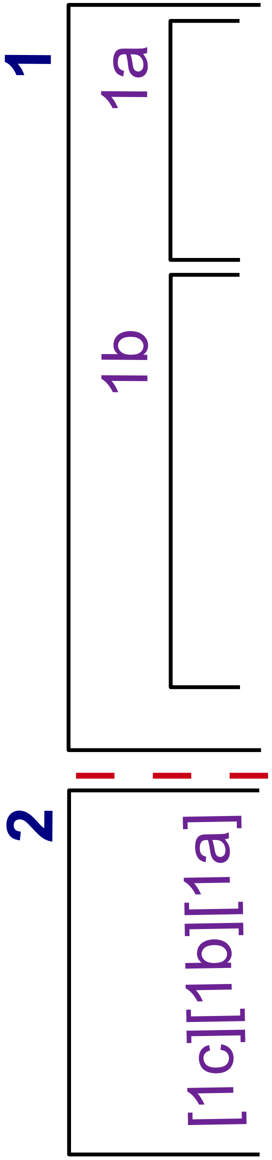
- **Approach 2**: Can we reduce the dimensionality of the data itself?
  - Two techniques:
    - Feature selection
      - Select features (dimensions) which look the most promising
      - Requires “useless” dimensions
    - Dimensionality reduction
      - Use statistical techniques to combine multiple dimensions into one
      - Linear projections → PCA, CSP, ICA, etc.
      - Nonlinear techniques → Sammon Mapping, SOM

(Recap) Course structure

Data Mining

1. Supervised
   a. Probability based
   b. Discriminant function

2. Unsupervised
   a. Clustering
   b. Visualization, dimensionality reduction
   c. Collaborative filtering



(1)Prediction and Classification with k-Nearest Neighbors

(2)Classification and Bayes Rule, Naïve Bayes

(3)Decision Trees

(4)Decision Trees II

(5)Discriminant Analysis

(6)Logistic Regression Case

(7)Neural Networks

(8)Neural Networks II

(9)Regularization and training algorithms.

(10)k-Means clustering, hierarchical clustering

(11)Visualization, Principal Components

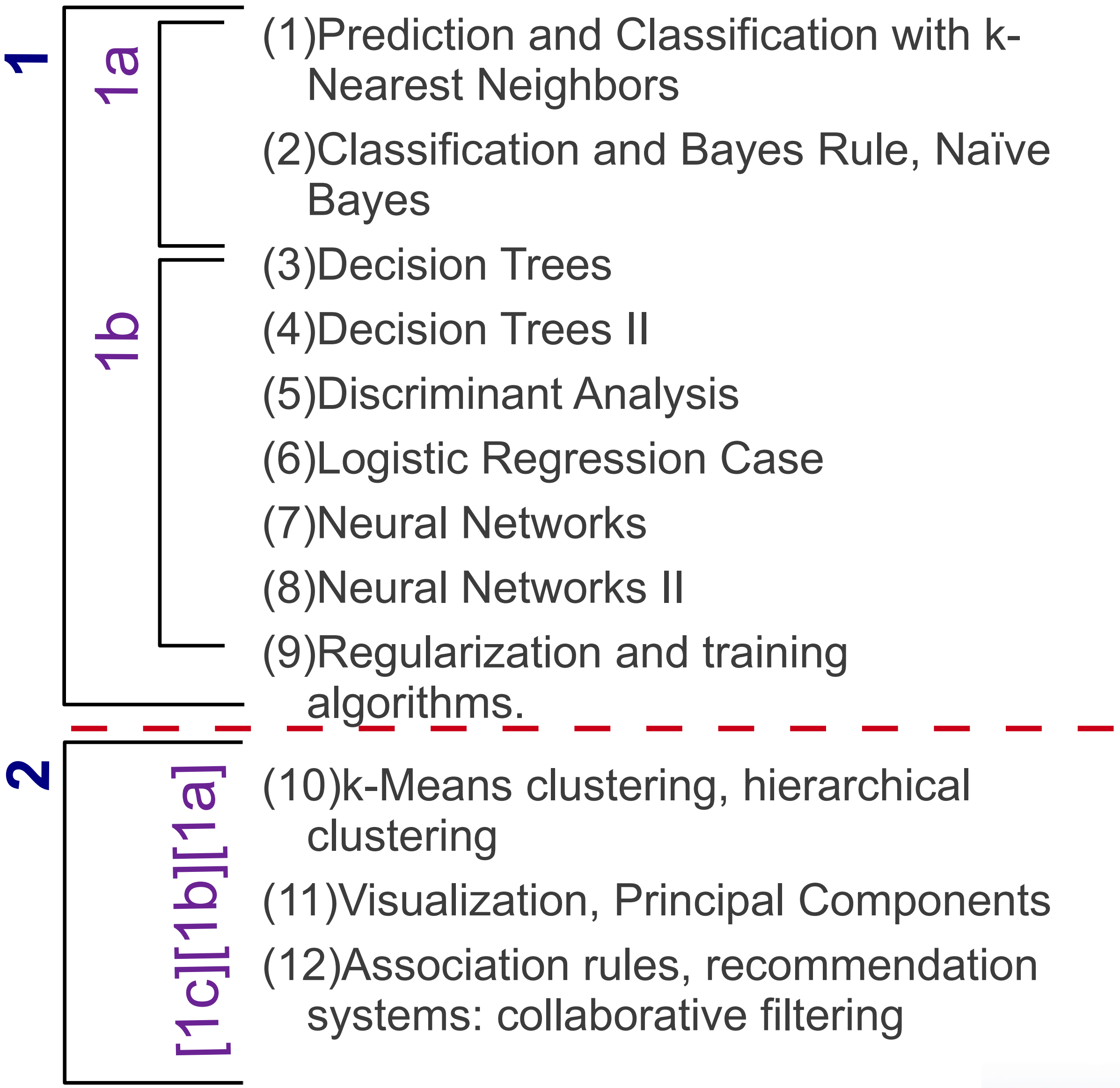
(12)Association rules, recommendation systems: collaborative filtering

(Recap) Course structure

Data Mining

1. Supervised
   a. Probability based
   b. Discriminant function

2. Unsupervised
   a. Clustering
   b. Visualization, dimensionality reduction
   c. Collaborative filtering



(1) Prediction and Classification with k-Nearest Neighbors

(2) Classification and Bayes Rule, Naïve Bayes

(3) Decision Trees

(4) Decision Trees II

(5) Discriminant Analysis

(6) Logistic Regression Case

(7) Neural Networks

(8) Neural Networks II

(9) Regularization and training algorithms.

(10) k-Means clustering, hierarchical clustering

(11) Visualization, Principal Components

(12) Association rules, recommendation systems: collaborative filtering

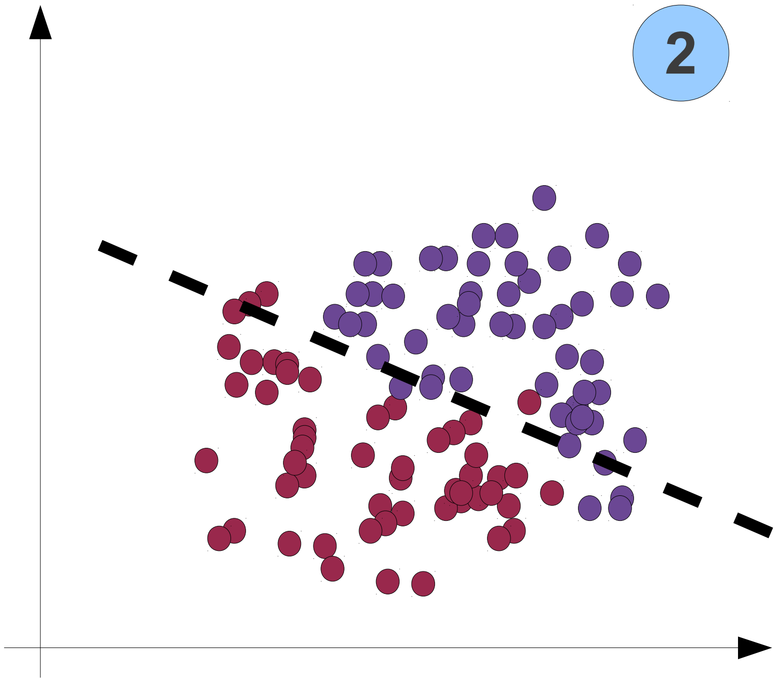
Types of data mining algorithms

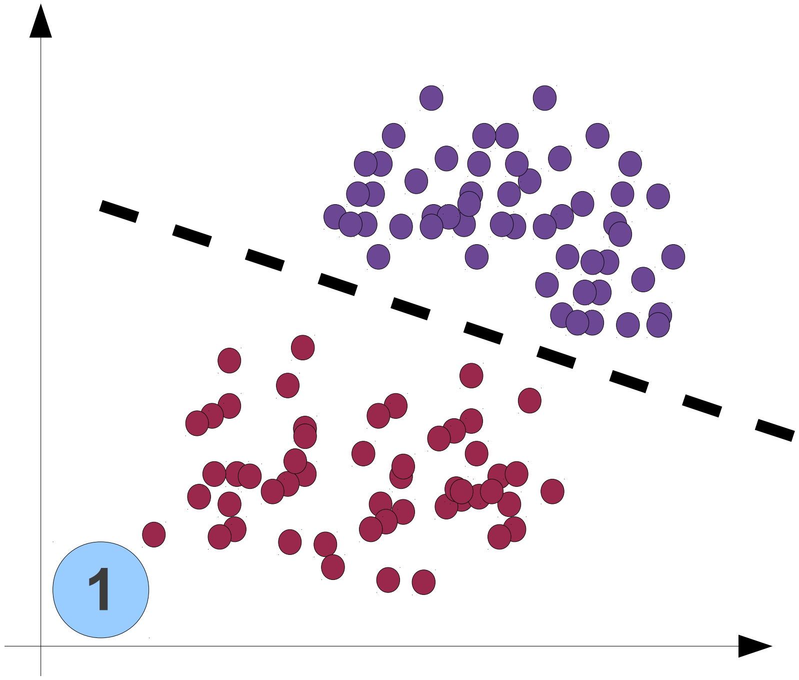
- **Two big classes**
  - *Supervised* learning algorithms
    - discover patterns in the data that relate data attributes with a target/class attribute.

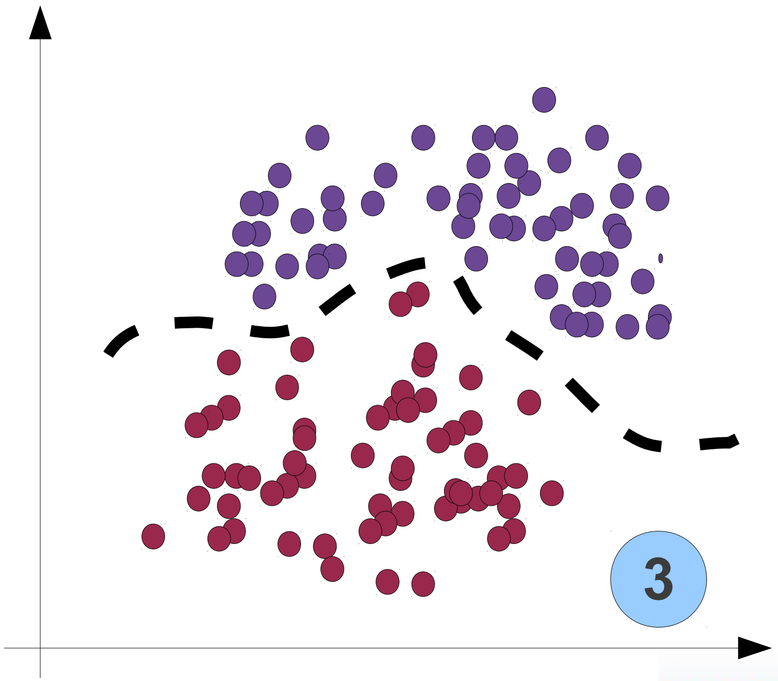
      (function approximation)
    - These patterns are then utilized to predict the values of the target attribute in future data instances.
    - Two broad cases:
      - Target attribute is numerical – known as "***Regression"***
      - Target attribute is categorical - "***Classification***"

(Classification)

1) Linear Classification (idealized)
2) Linear Classification (real world)
3) Non-linear Classification







● Class A

● Class B

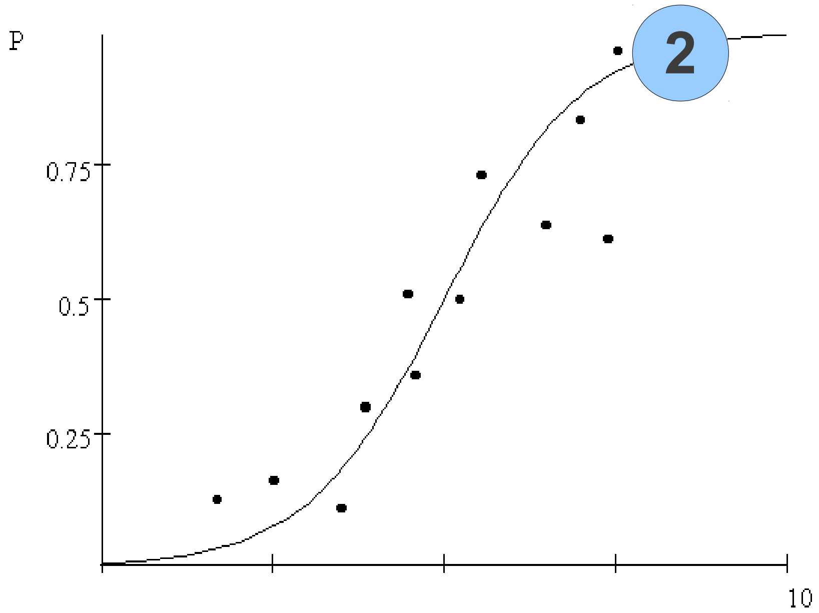
(Regression)

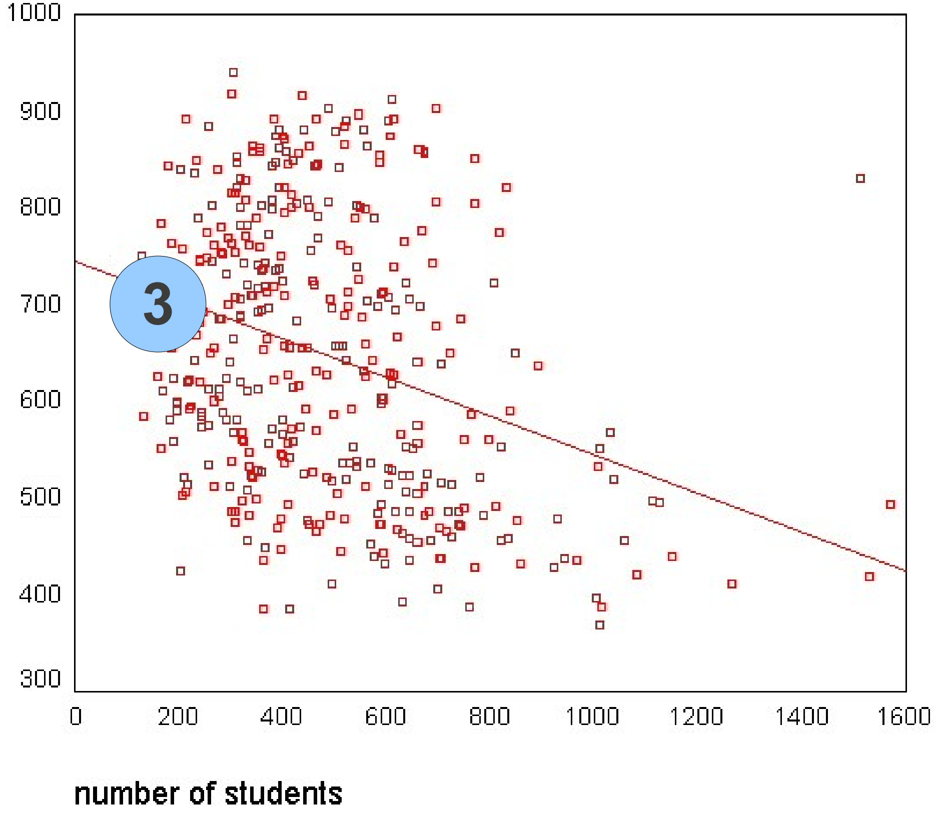
(1) Linear Regression

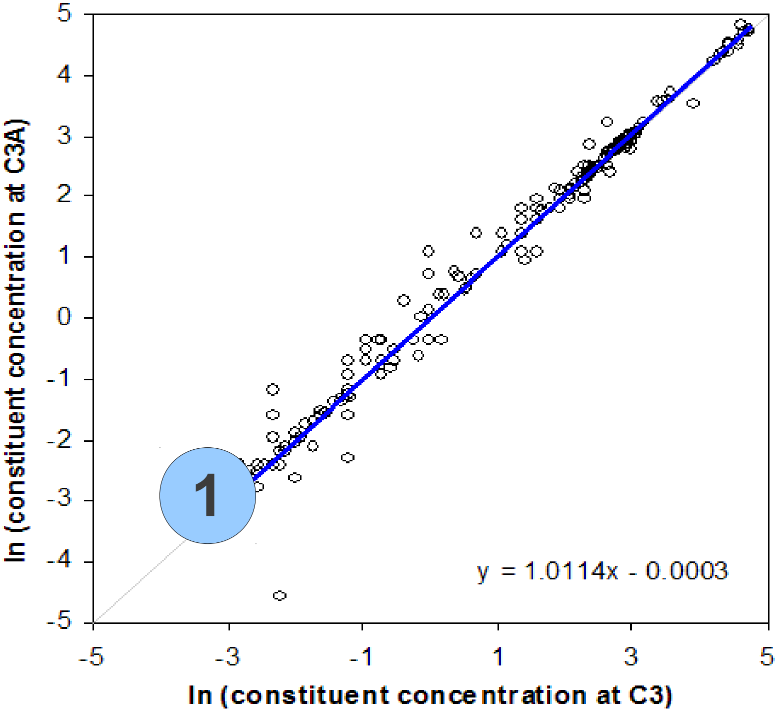
(2) Logistic Regression

(3) (Very bad!) Linear Regression

Not covered in course





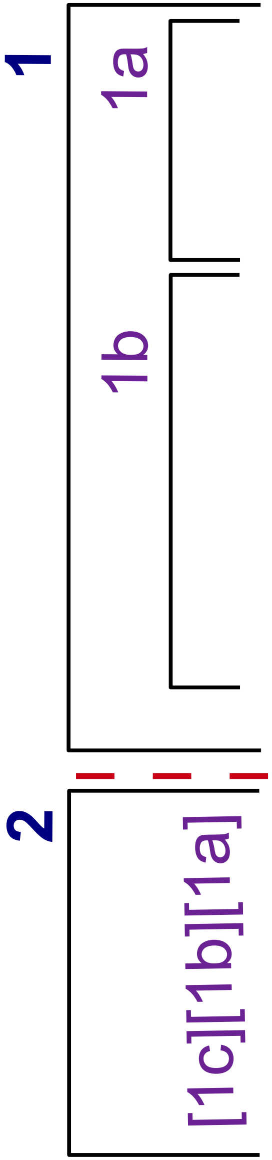


(Recap) Course structure

Data Mining

1. Supervised
   a. Probability based
   b. Discriminant function

2. Unsupervised
   a. Clustering
   b. Visualization, dimensionality reduction
   c. Collaborative filtering



(1)Prediction and Classification with k-Nearest Neighbors

(2)Classification and Bayes Rule, Naïve Bayes

(3)Decision Trees

(4)Decision Trees II

(5)Discriminant Analysis

(6)Logistic Regression Case

(7)Neural Networks

(8)Neural Networks II

(9)Regularization and training algorithms.

(10)k-Means clustering, hierarchical clustering

(11)Visualization, Principal Components

(12)Association rules, recommendation systems: collaborative filtering

(cont'd)

- **Unsupervised**
  - Data has no target attribute (or may choose not to use target attributes)
  - We want to explore the data to find some intrinsic structures in them.
  - ie: Supervised → “*here is the **answer***”

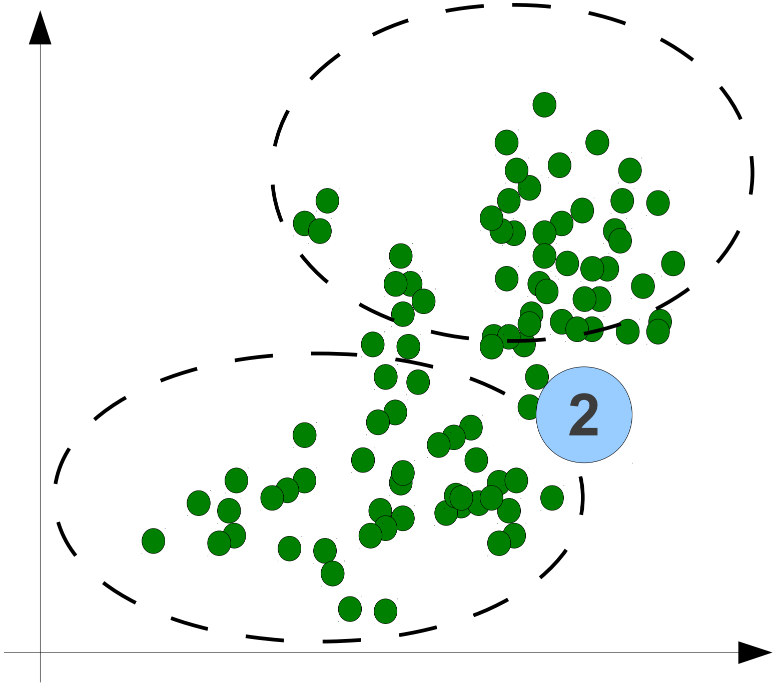
    Unsupervised → “*here is the **question***”
- **Examples:**
  - Clustering
  - Tree/Topology inference
  - Density estimation
  - “Maps”
  - Topographic projections

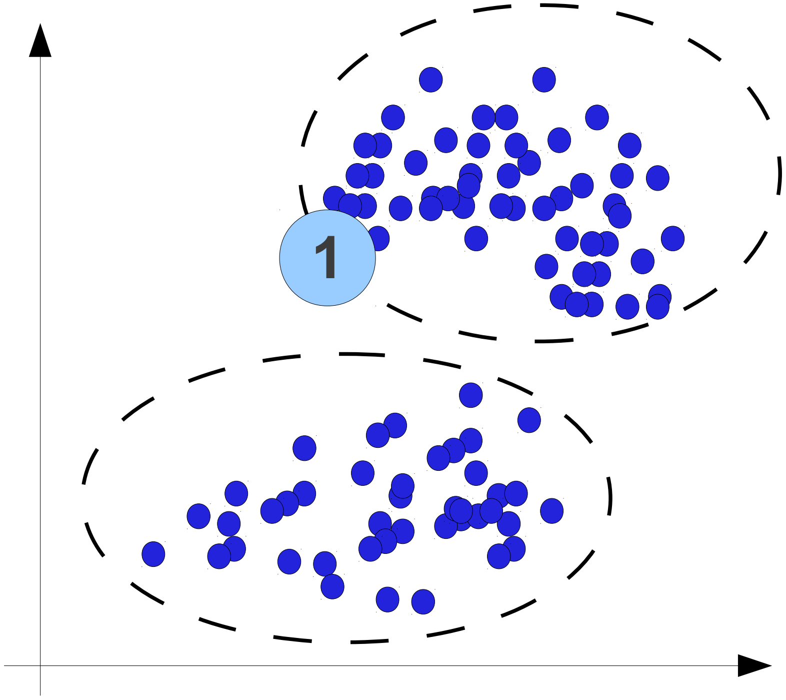
(Clustering)

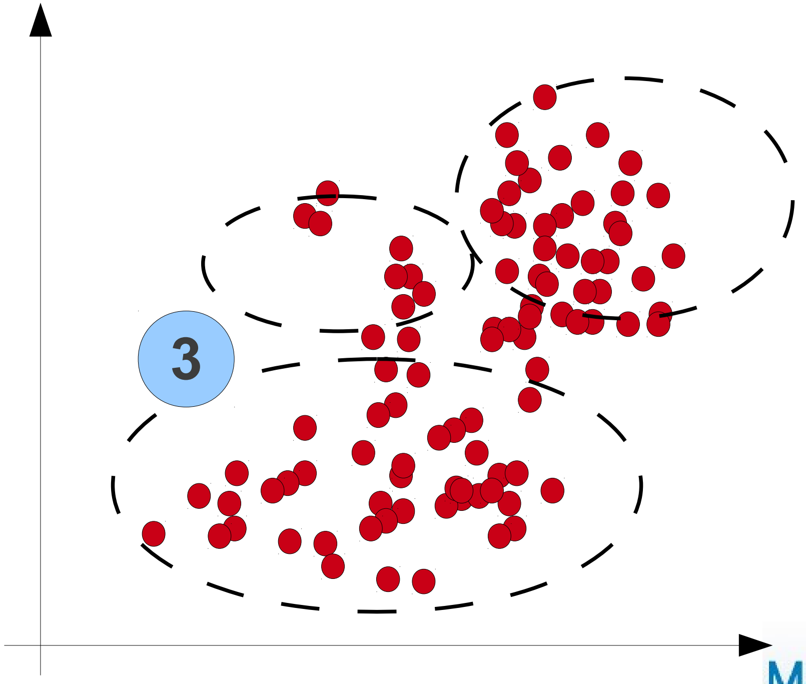
(1) 2 cluster case (fantasy land)

(2) 2 cluster case (real world)

(3) 3 cluster case (real world)







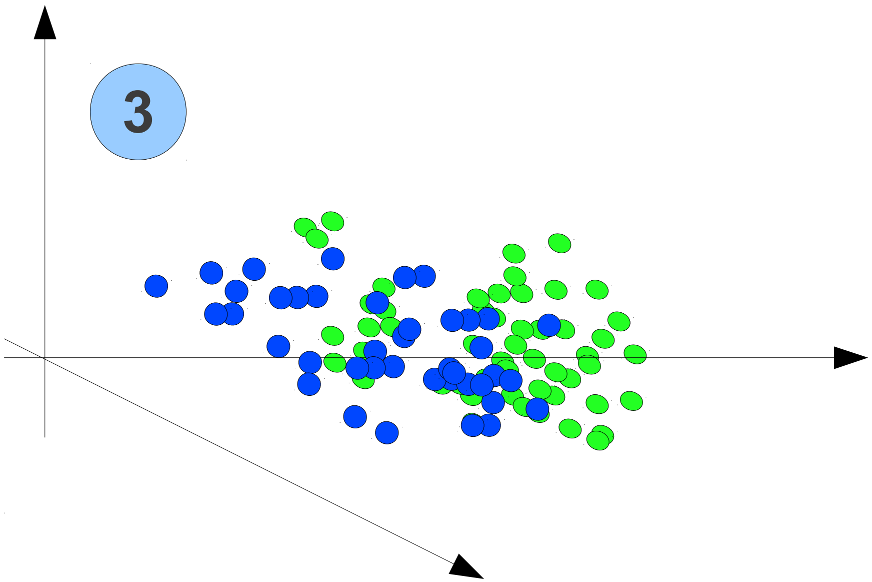
(Cont'd)

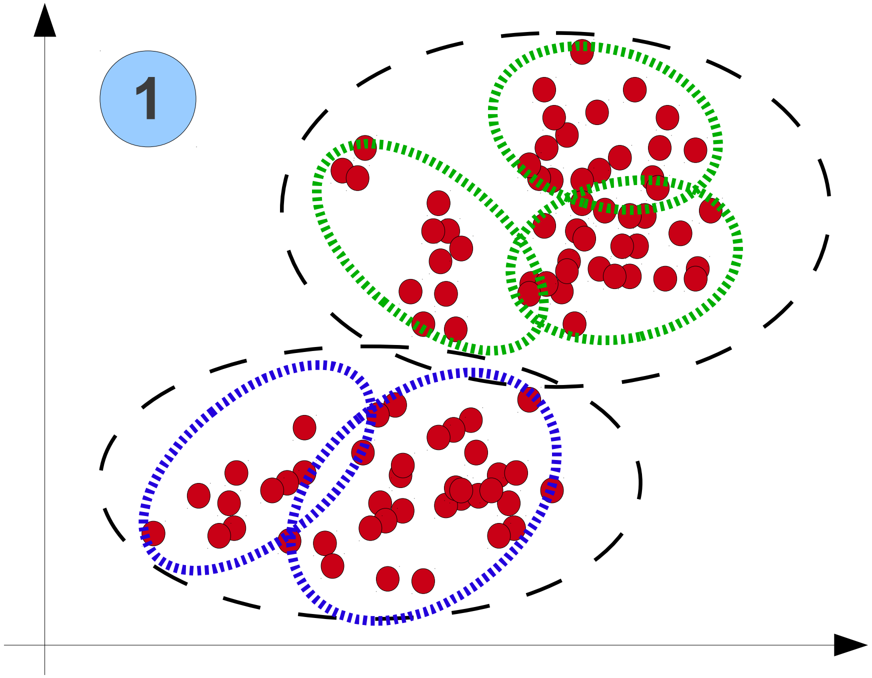
(1) Hierarchical clustering

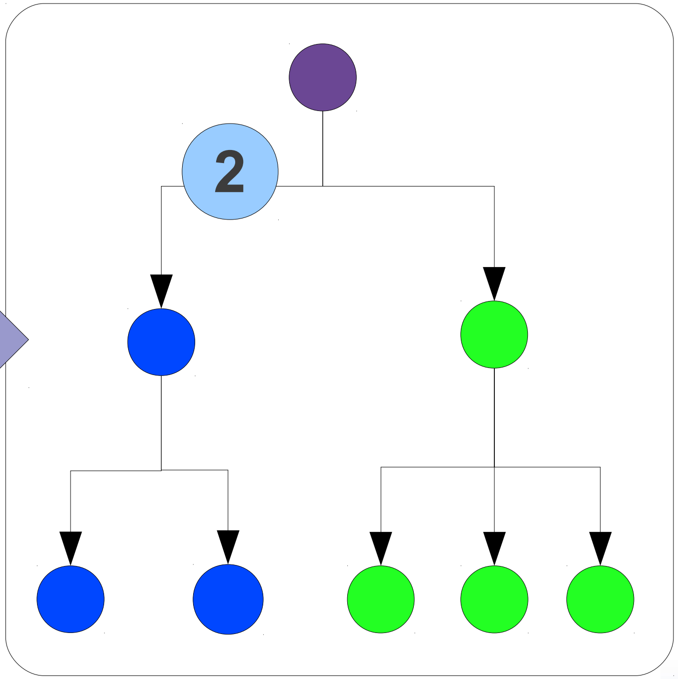
(2) Inference of taxonomy

(3) Topographic mapping

(***Visualization***)







Other topics (not /partially covered)

- **Alternative forms of learning**
  - Semi-supervised
  - Reinforcement
  - Transfer learning
- **Other “dimensions” or levels**
  - Optimization algorithms
  - Meta-learning
  - Feature extraction
  - Signal processing
  - Learning theory, etc..
  - Etc..!