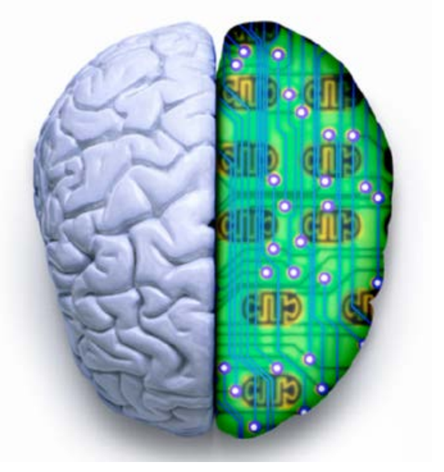
Advanced Artificial Intelligence CM4107 (Week 4)



Machine Learning

Dr. Yann Savoye
School of Computing Science and Digital Media
Robert Gordon University

Module Information

- **Assessment:**

  - Coursework (2 components)
    - Component 1: literature review
    - Component 2: paper implementation

  - No mid-term or final written exam.

    All deadlines are strong:
    - *It will not be possible to upload material after the deadline.*
    - *No deadline extension will be granted. No excuse.*
    - *Only the content submitted via the Moodle will be mark.*

Coursework

- **Submission of the Coursework Part 1:**

  Deadline - Monday, October 30th, 2017 23:00:

  Activity 1 and Activity 2

  - *2-pages written reports (in PDF format)*
  - *7 slides presentation (in PDF format)*

- You should have read two papers by now ! And started to sketch a draft for the 2-pages report …

Coursework

- **How to write a critical review?**

"Convergence and Divergence"

Convergence

Divergence

Overview

- Part I – Machine Learning
- Part II – Classical Problems
- Part III – Linear Model
- Part IV – K-Nearest Neighbours
- Part V – K-Means
- Part VI – Principle Analysis Component
- Part VI – Dimensionality Reduction
- Part VII – Decision Tree and Forest
- Part VIII – Principal Component Analysis

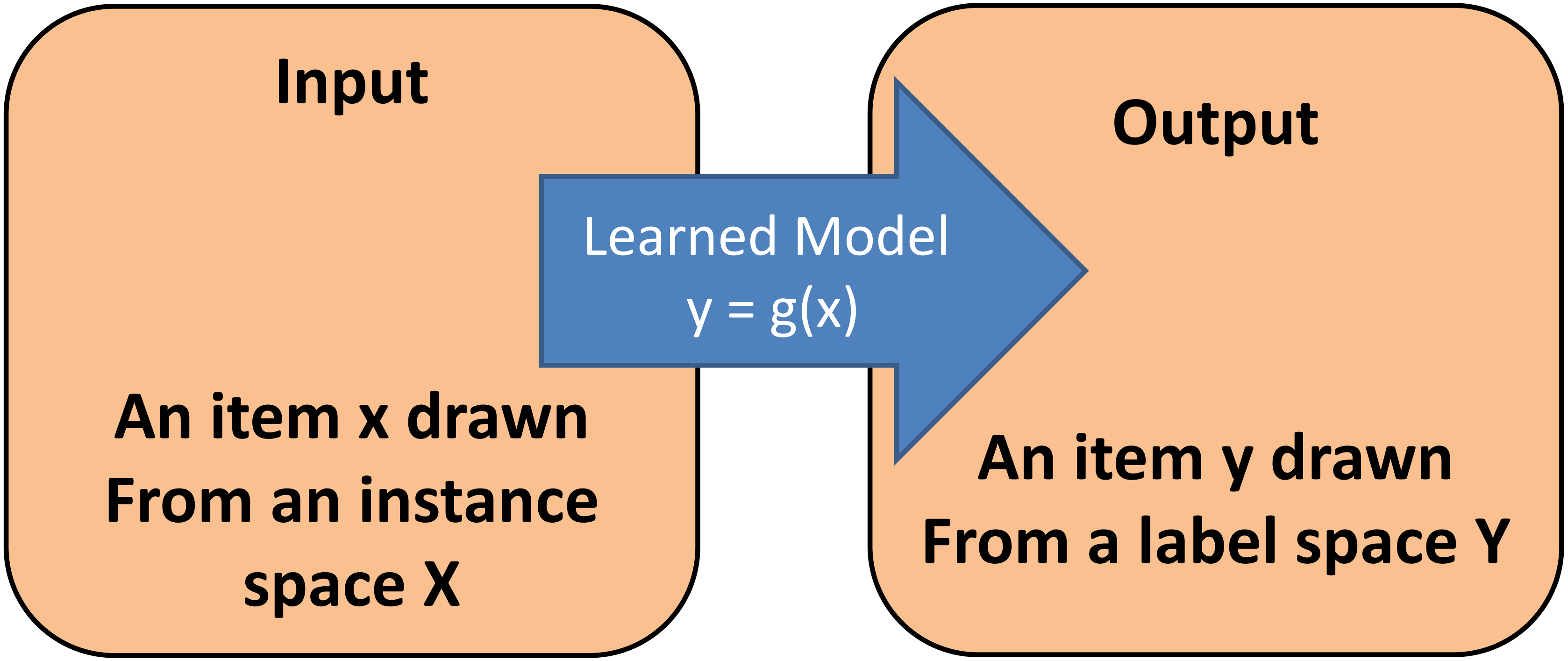
Part I – Machine Learning

Self-Learning Program

Arthur Samuel, The World’s first Self Learning Program

*In 1956 Samuel showed the **capability of the computer** on television by demonstrating his **checkers learning program**, widely regarded to be the **world’s first self-learning program**.*

Machine Learning



“You need to know the question you are trying to answer”

- Jason Bell (2015)

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed“

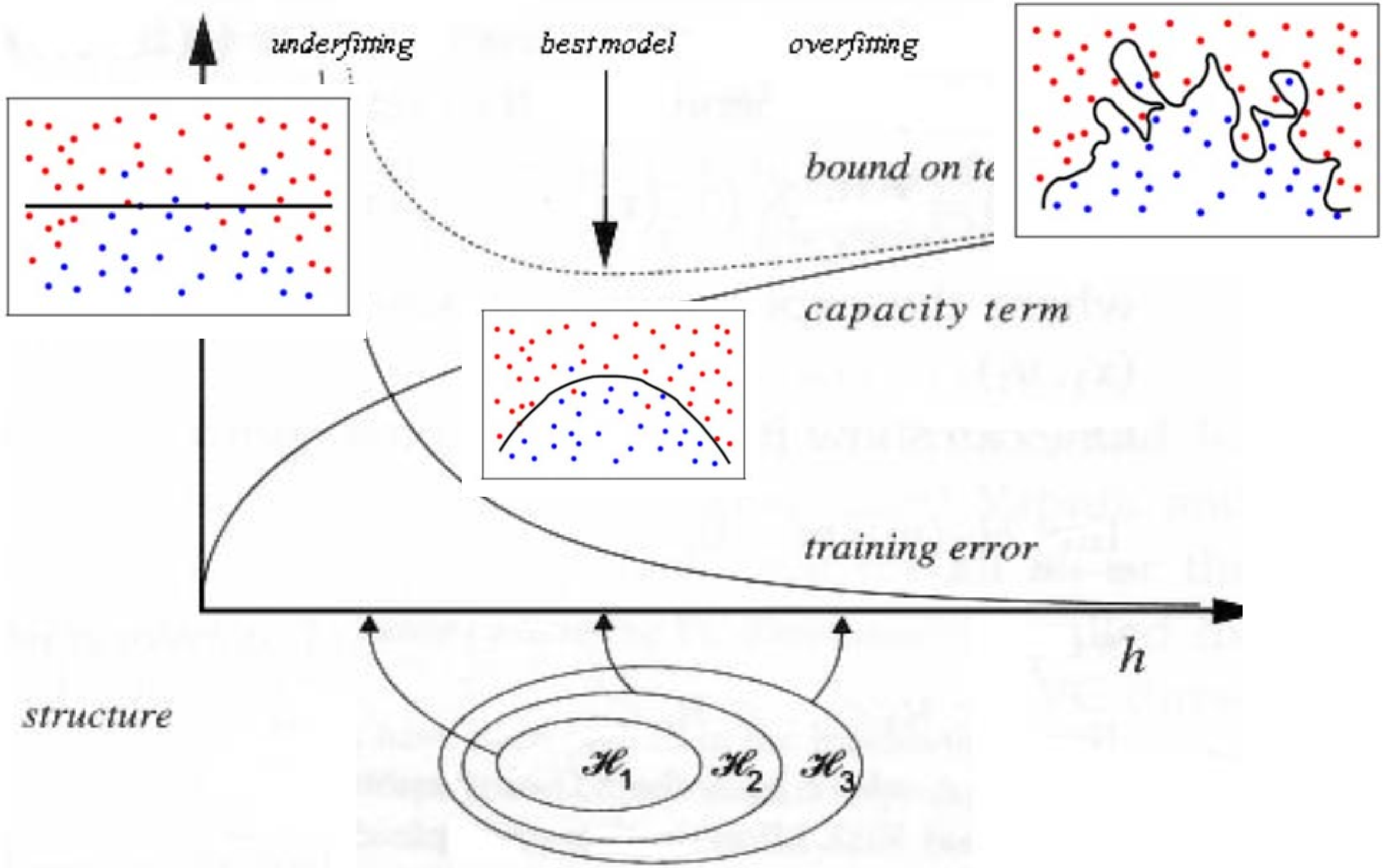
- Machine learning is the science of getting computers to act **without being explicitly programmed**.

- The best way to make progress towards **human-level AI.**

- **Machine learning** is employed in a range of computing tasks where designing and programming explicit algorithms with good **performance is difficult or infeasible.**

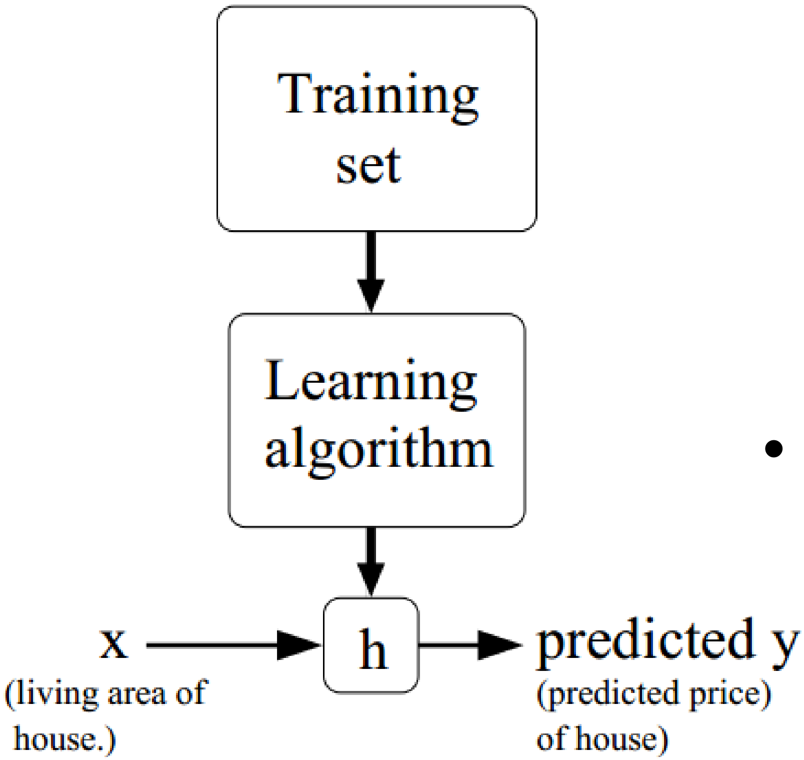
- Effective machine learning is difficult because **finding patterns is hard** and often not enough training data is available.

Importance of the Model



*Constructing **models** that predict **data distributions**.*

Supervised or Unsupervised?



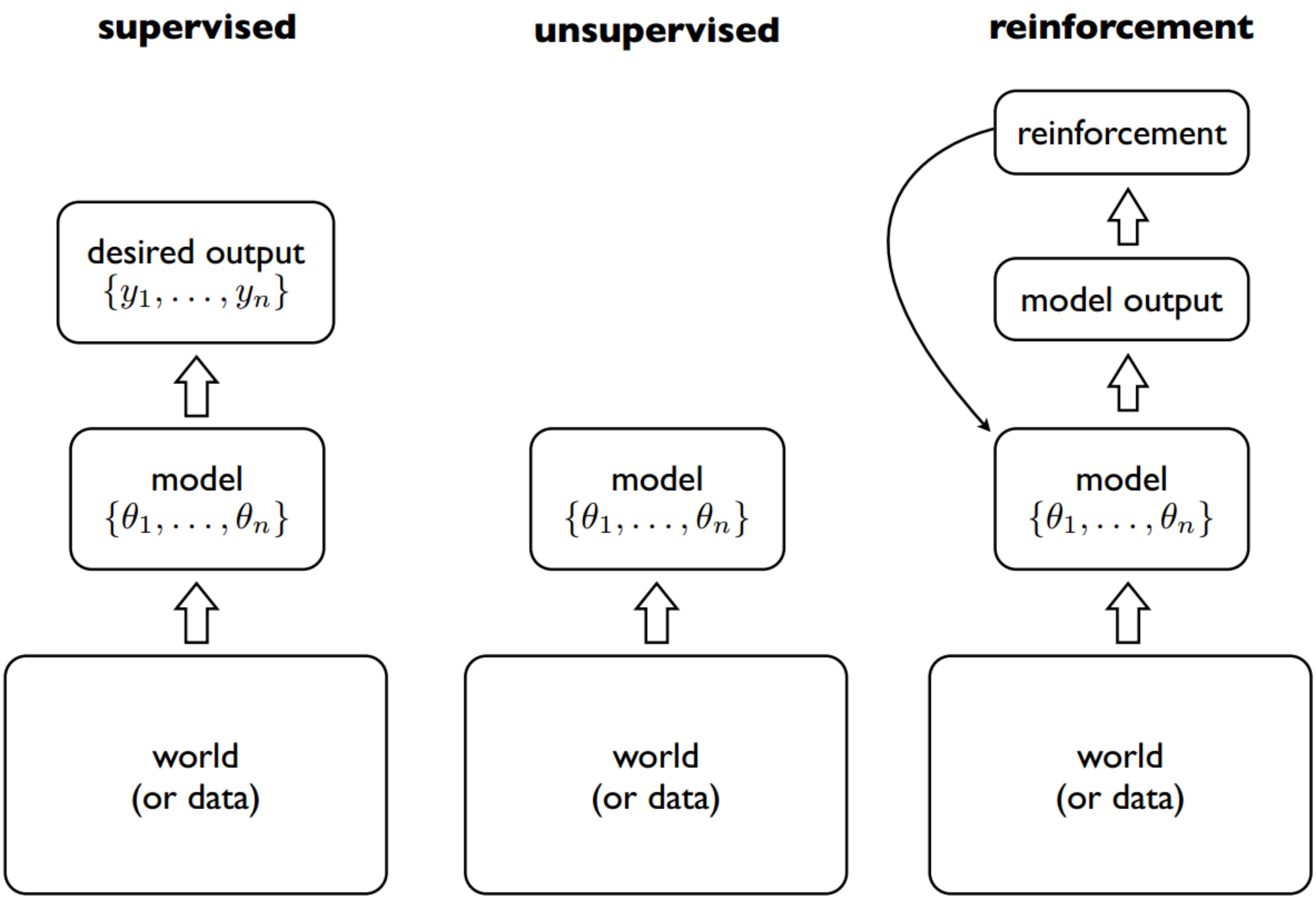
- **Supervised Learning**: given a sample of input-output pairs (training samples), then the task is to find a deterministic function that maps any input to an output that can predict future input-output observation.

  => labelled samples

- **Unsupervised Learning:** In unsupervised learning data samples are given without target values, Typical example of unsupervised learning include text and image segmentation and analysis.

  => unlabelled samples

Types of Learning



Machine Learning

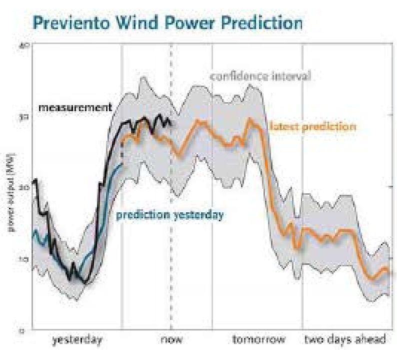
- **Linear Regression**
- **Support Vector Machine**
- **Principal Component Analysis**
- **K-Mean**
- **Random Forest**

- Expectation Maximization
- Bayesian Network
- Neural Network
- Deep Learning

Machine Learning

Machine learning extracts features from data to solve many different **predictive tasks:**

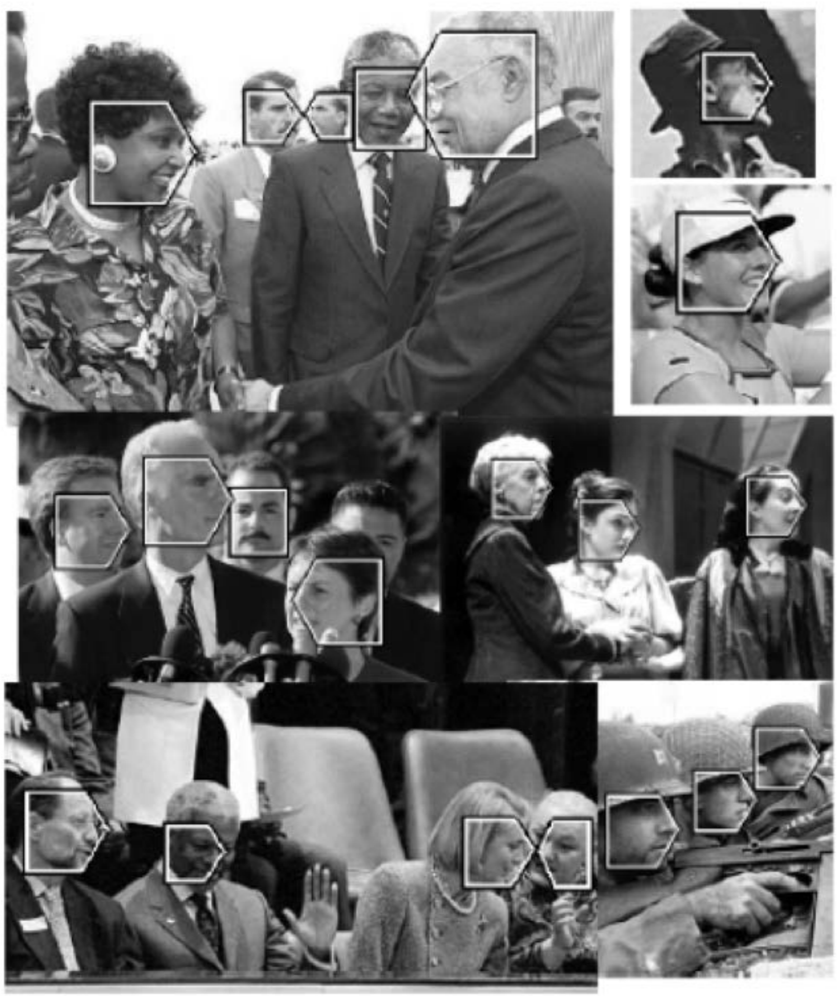
- Forecasting (energy, sales)
- Imputing missing data
- Detecting anomalies
- Classifying
- Ranking
- Summarizing
- Decision making.

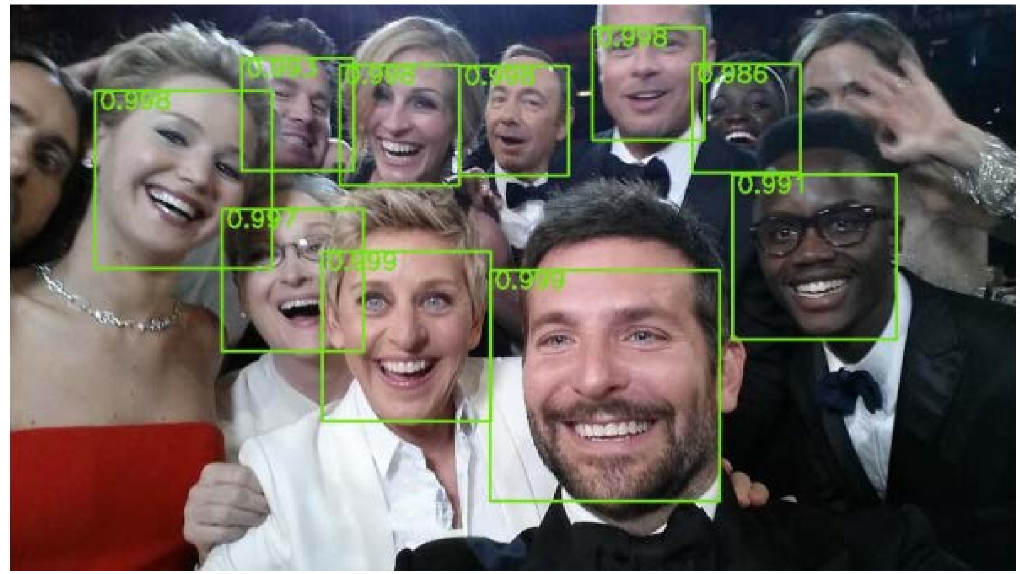


Part II – Classical Problems

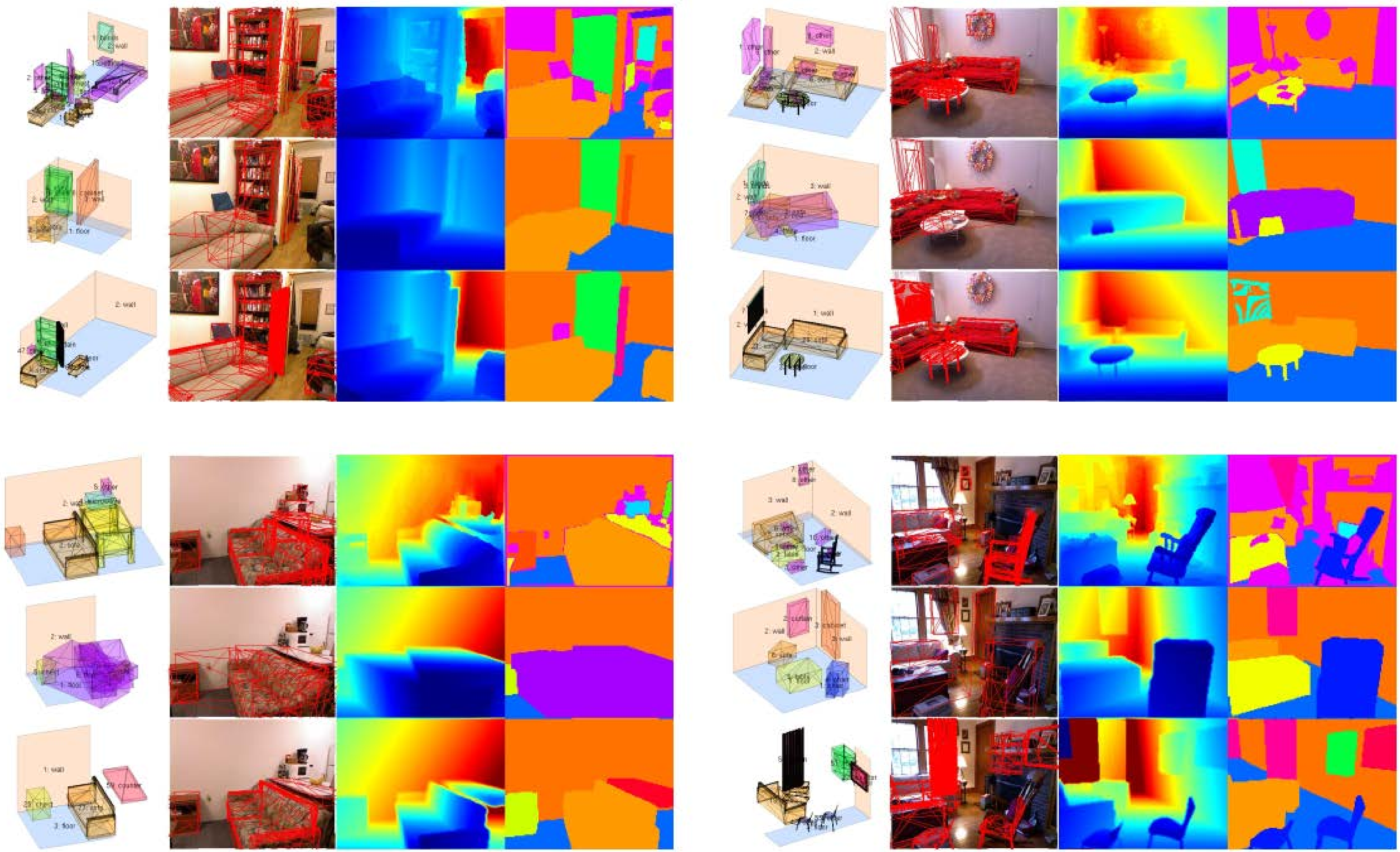
Object Detection

Face Detection

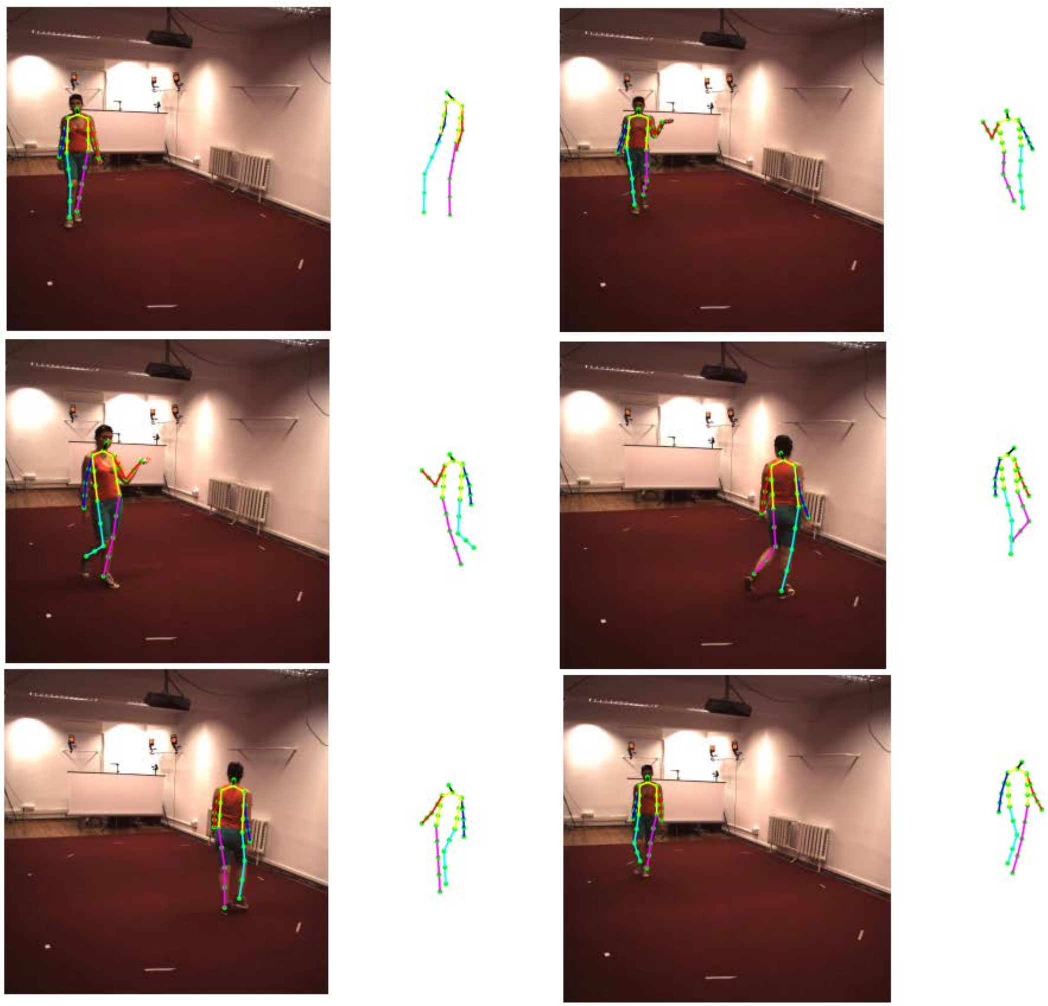


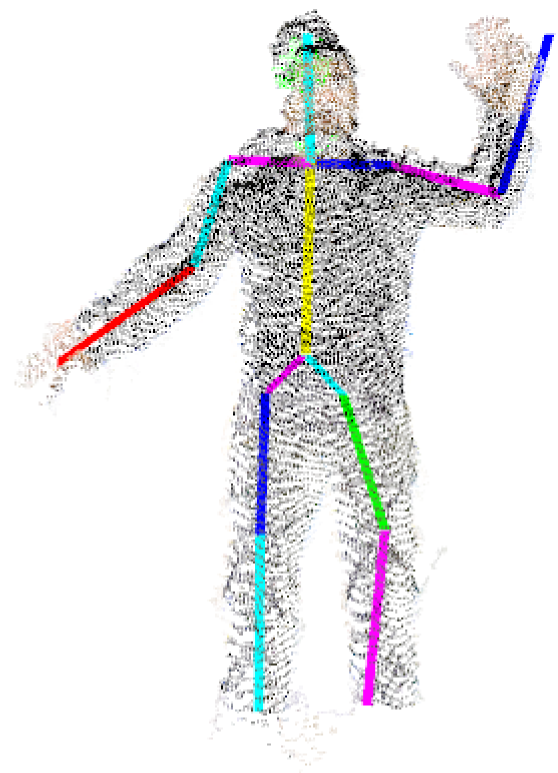


Scene Understanding

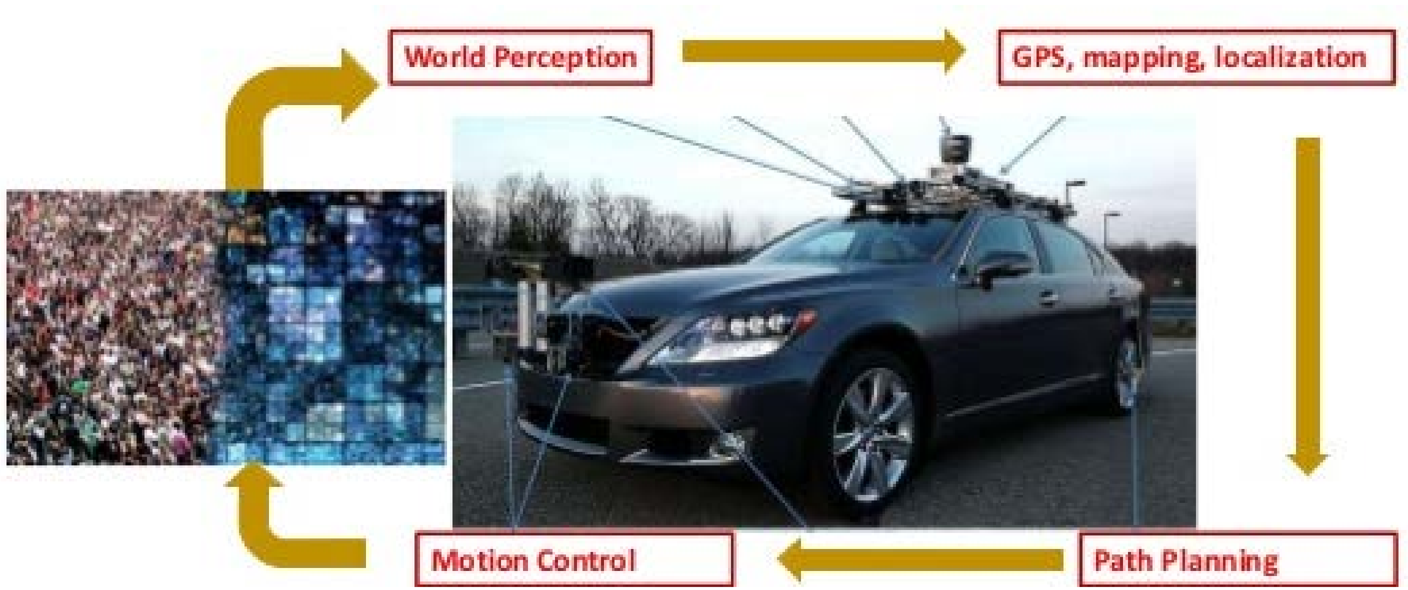


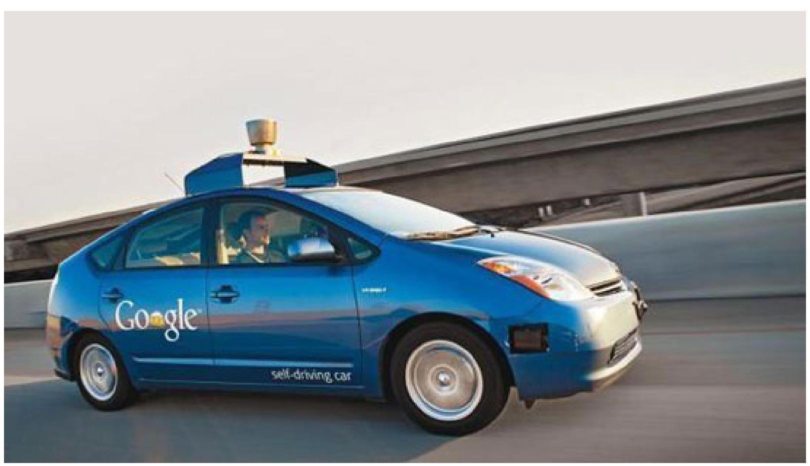
Pose Estimation





Computer Vision and Robotics



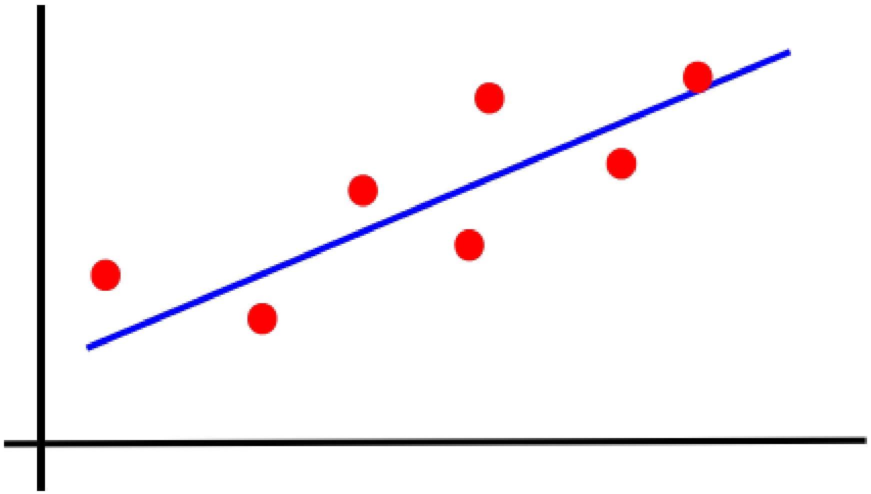


Classification

Part III – Linear Model

Linear Regression

x	y
1.0	2.6
2.3	2.8
3.1	3.1
4.8	4.7
5.6	5.1
6.3	5.3

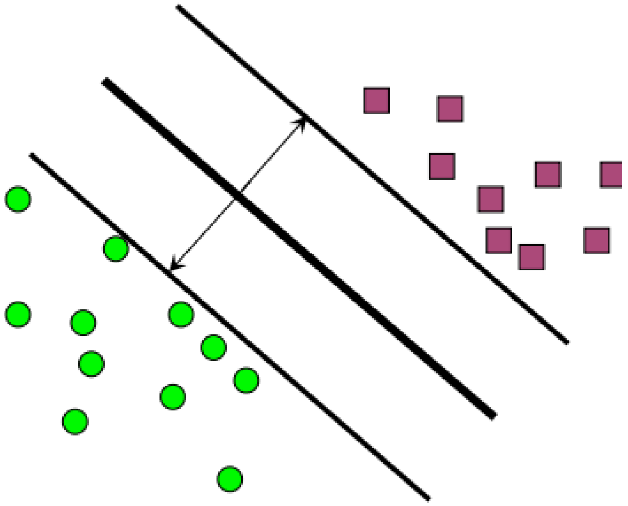


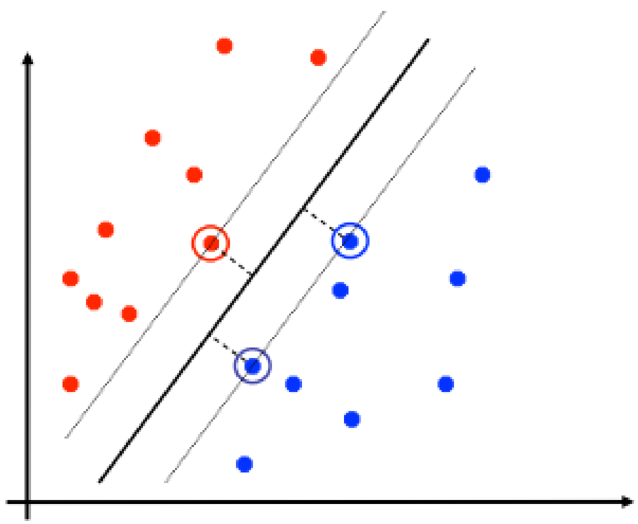
- **Linear regression**, the line/plane/hyperplane that we found was calculated to be as closes as possible to all the data points.

- In **regression**, the output space is formed by the values of continuous variables. Typical example of regression is to predict the value of shares in the stock exchange market.

Support Vector Machine

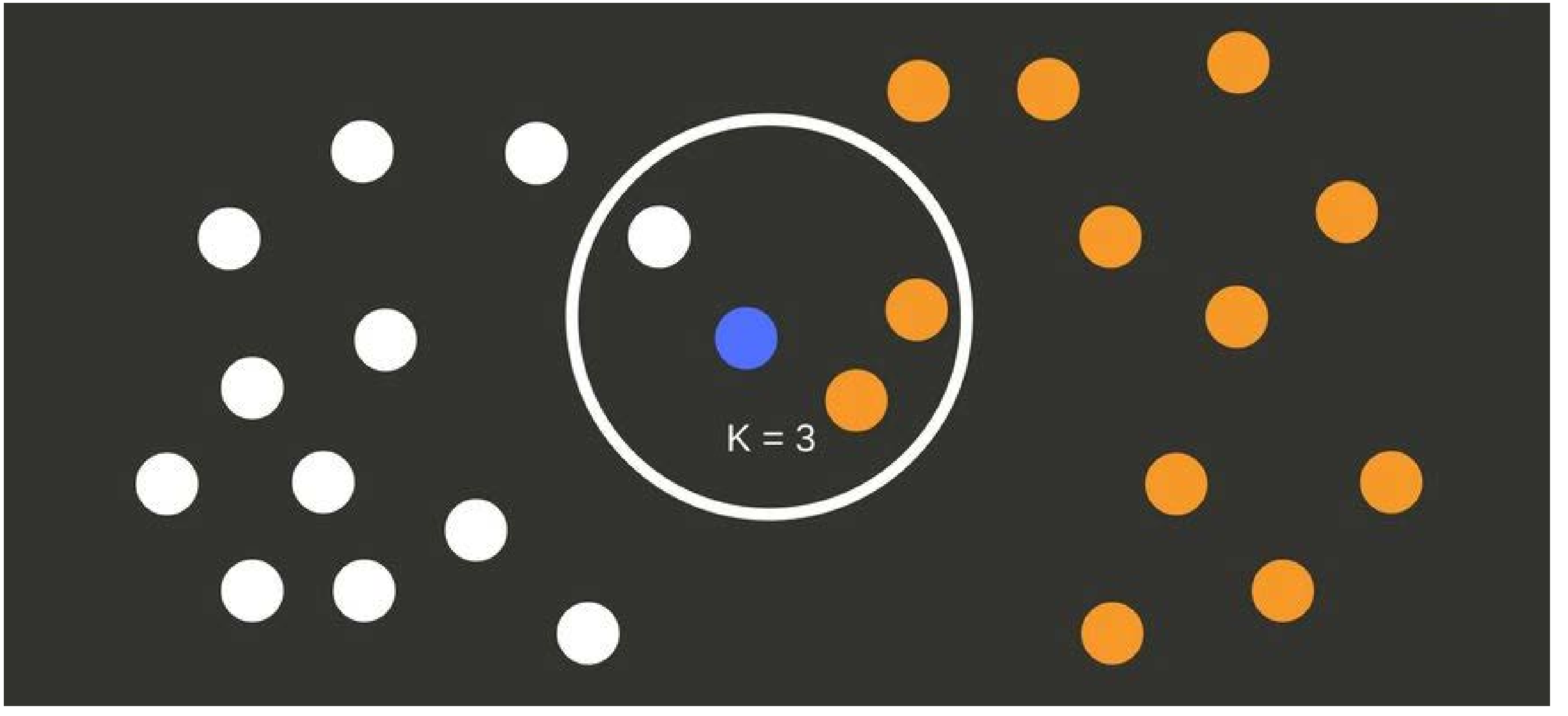
- **Replace the original data** with a simplified (linear) model.
- The input should **be linearly separable**.
- **Hyperplane** to represent the border between the two differently colored regions.
- Determined by only **the few data points that are closest to the border** between the two classes.
- Move the line to a position that **minimizes the thickness of the band**.





Part IV – K-Nearest Neighbours

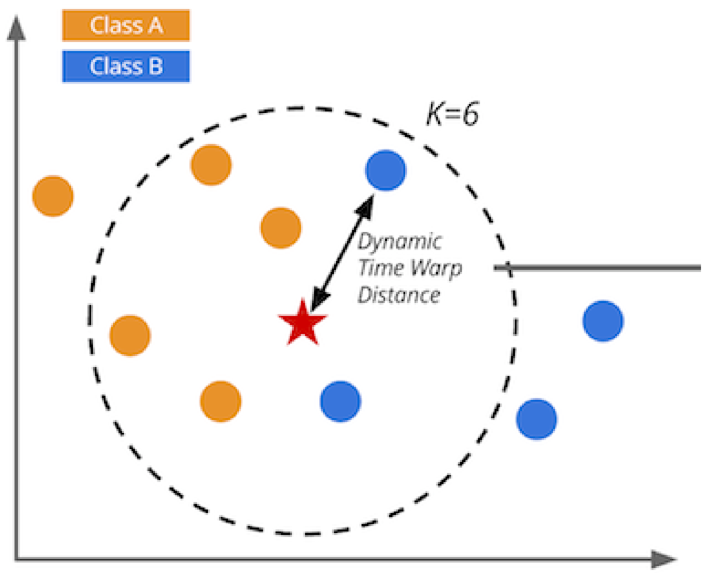
K-Nearest Neighbours

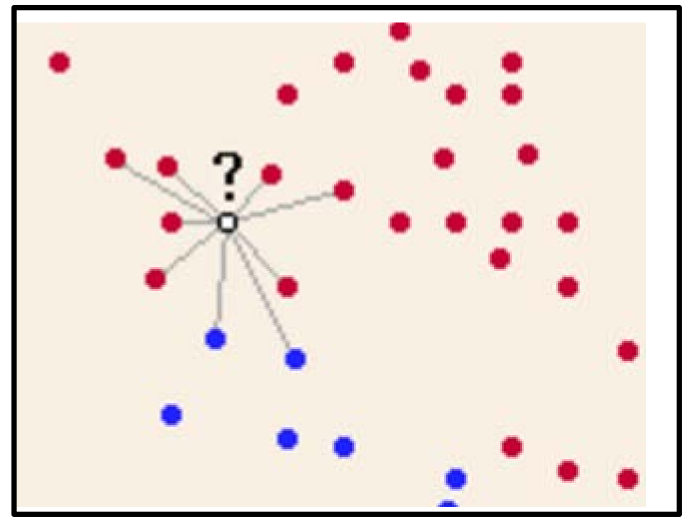


(image courtesy of dataaspirant.com)

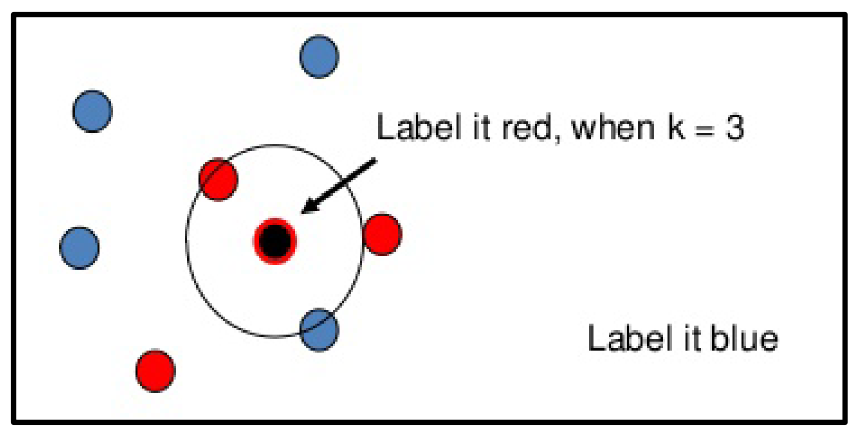
*The **k-nearest neighbors** algorithm is one of the most basic yet essential supervised learning algorithm.*

K-Nearest Neighbors



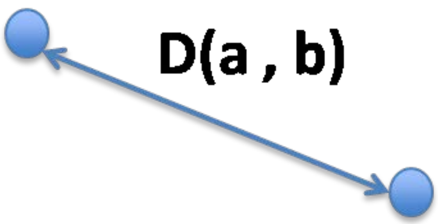


*the **k-nearest neighbors** algorithm (k-NN) is a non-parametric method*
*Used for **classification** and **regression***

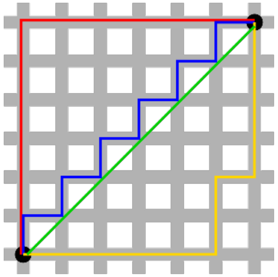


Distance Measures

- Euclidean: $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

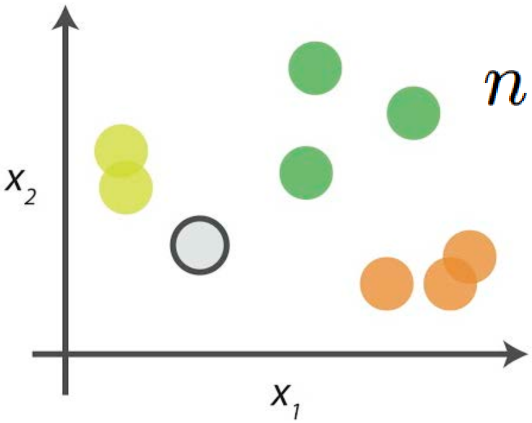


- Manhattan: $\sum_{i=1}^{k}|x_i - y_i|$

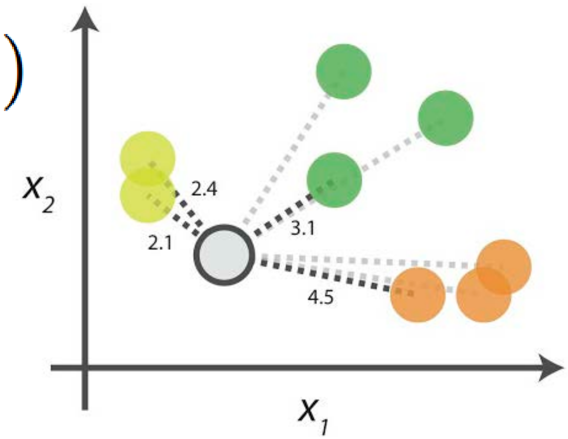


- Minkowski: $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

K-Nearest Neighbours



$$n^* = \arg\min_{n \in D} dist(x, x_n)$$



Algorithm 1.2 k-Nearest Neighbor Classification

Classify($\mathbf{X}, \mathbf{Y}, x$) {reads documents $\mathbf{X}$, labels $\mathbf{Y}$ and query x}

  for $i = 1$ **to** m **do**

    Compute distance $d(x_i, x)$

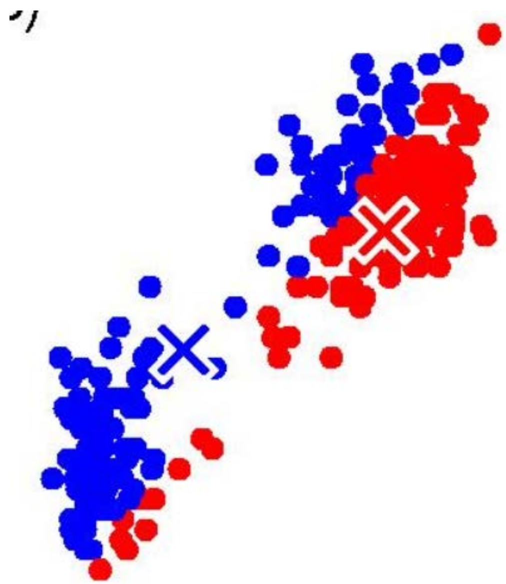
  end for

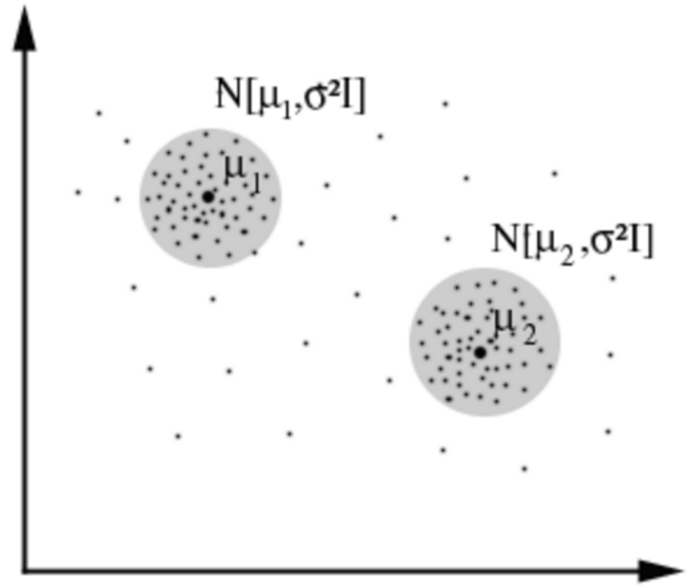
  Compute set I containing indices for the k smallest distances $d(x_i, x)$.

  return majority label of $\{y_i \text{ where } i \in I\}$.

Part V – K-Means Clustering

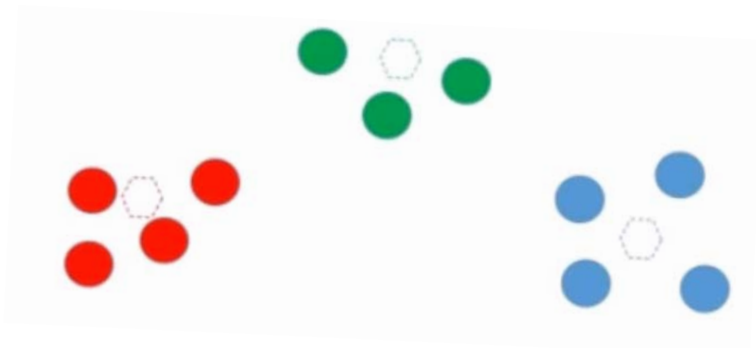
K-Mean Clustering

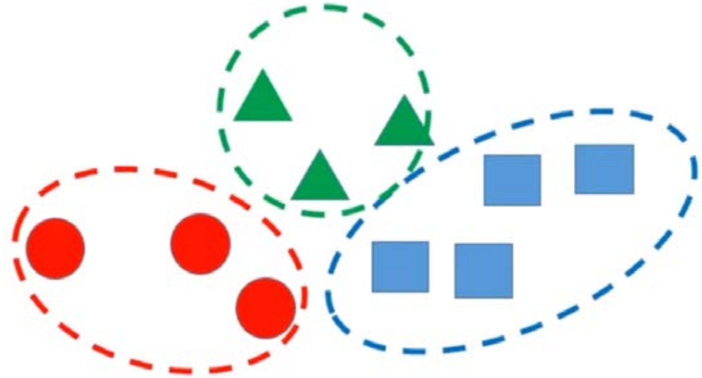




- The process of **organizing objects into groups.**
- **A *cluster*** is a collection of objects which are "**similar**" between them and are "**dissimilar**" to the objects belonging to other clusters.
- **K-means** clustering classifies objects based on features into K groups.
- The grouping is done by **minimizing the sum of squares of distances** between data and the corresponding cluster centroid.

Clustering





number of clusters

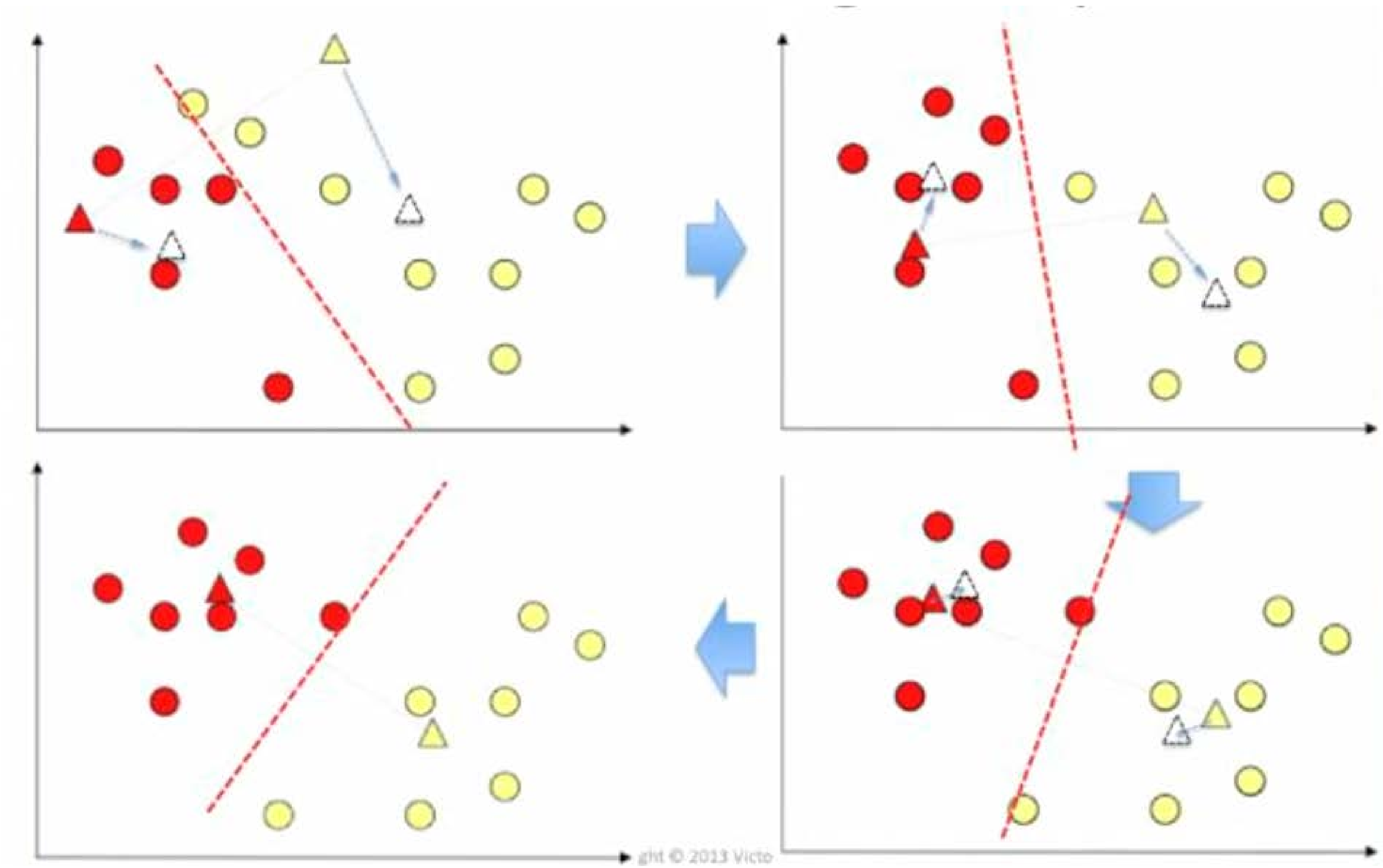
number of cases

centroid for cluster *j*

case *i*

$$\leftarrow J = \sum_{j=1}^{k}\sum_{i=1}^{n} \underbrace{\left\| x_i^{(j)} - c_j \right\|^2}$$

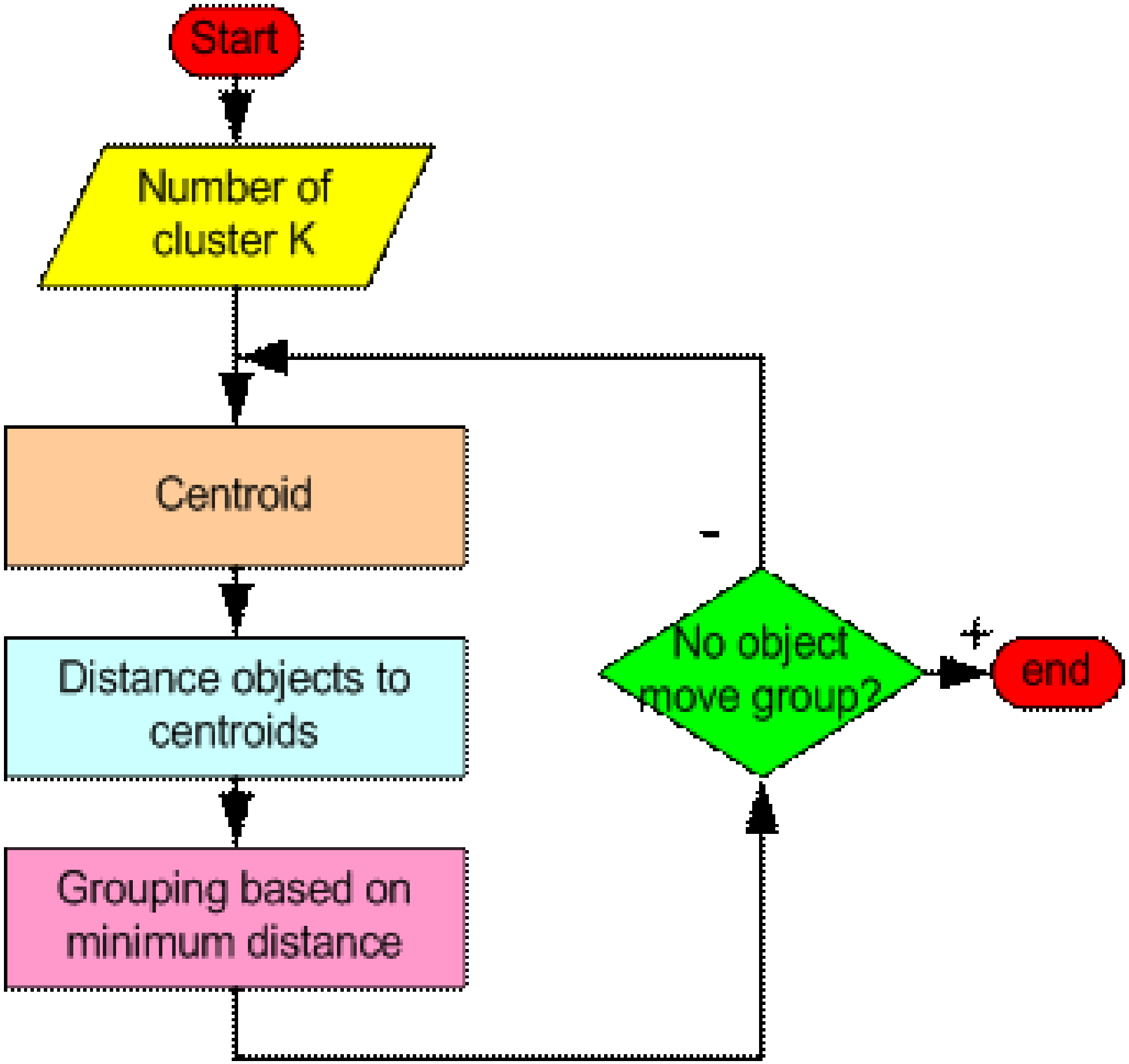
K-Means Clustering



K-Mean Clustering

An algorithm to automatically group the data into coherent clusters.

K-means *is by far the most widely used* ***clustering algorithm***



K-Means Pseudo-Code

Algorithm 1.5 K-Means

Cluster($\mathbf{X}$) {Cluster dataset $\mathbf{X}$}
    Initialize cluster centers μ_j for $j = 1, \ldots, k$ randomly
    repeat
        for $i = 1$ **to** m **do**
            Compute $j' = \operatorname{argmin}_{j=1,\ldots,k} d(x_i, \mu_j)$
            Set $r_{ij'} = 1$ and $r_{ij} = 0$ for all $j' \neq j$
        end for
        for $j = 1$ **to** k **do**
            Compute $\mu_j = \frac{\sum_i r_{ij} x_i}{\sum_i r_{ij}}$
        end for
    until Cluster assignments r_{ij} are unchanged
    return $\{\mu_1, \ldots, \mu_k\}$ and r_{ij}

K-Means Algorithm

1. Initialize **cluster centroids** $\mu_1, \mu_2, \ldots, \mu_k \in \mathbb{R}^n$ randomly.

2. Repeat until convergence: {

   For every i, set

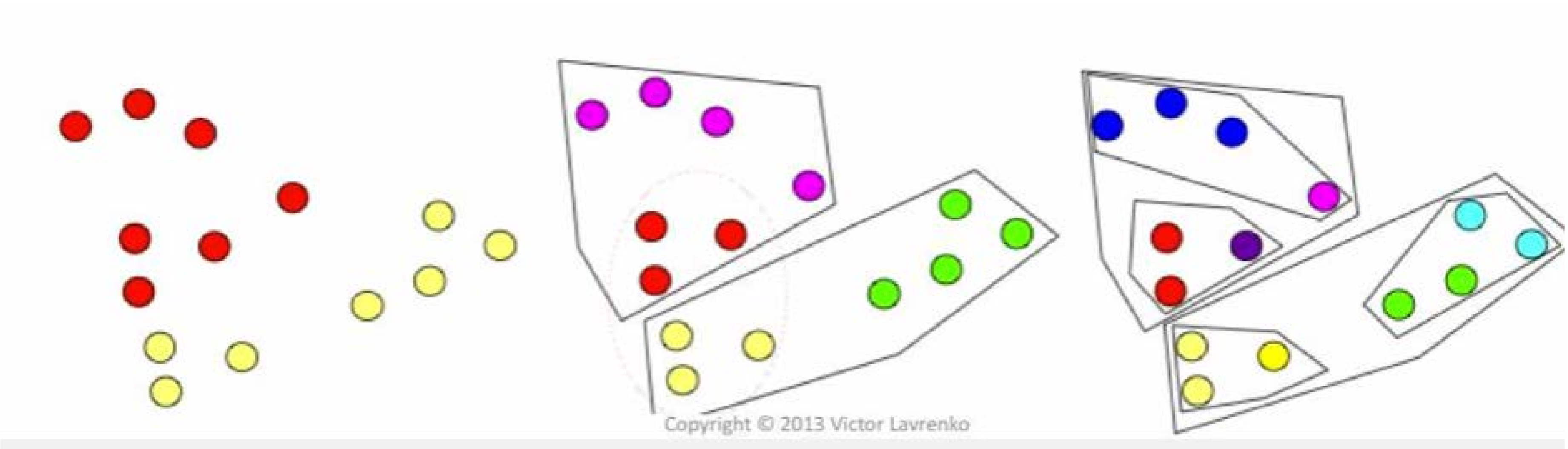
$$c^{(i)} := \arg\min_j ||x^{(i)} - \mu_j||^2.$$

   For each j, set

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}.$$

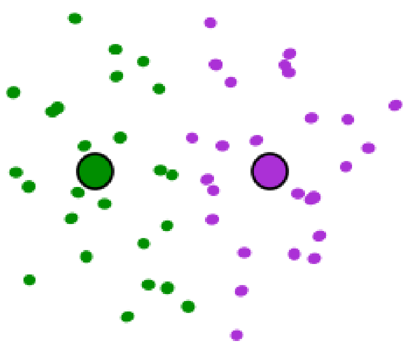
   }

Hierarchical K-Means

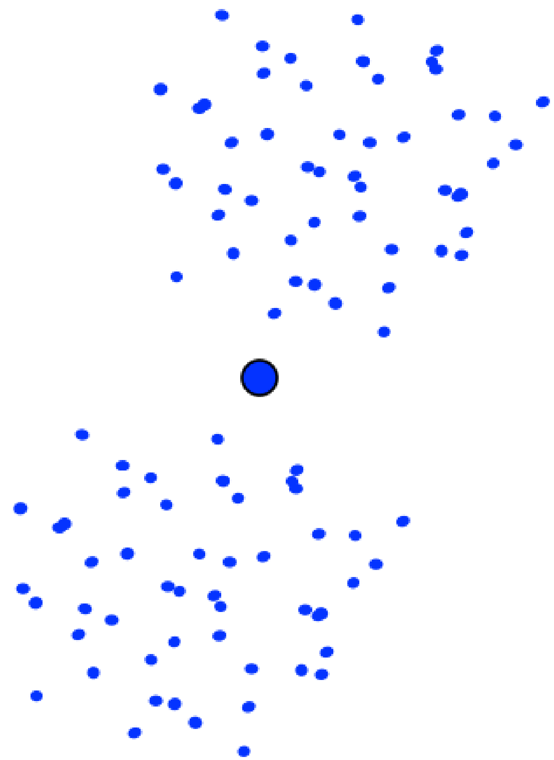


***Determining** the **number of clusters** is a challenging task.*

K-Means Getting Stuck



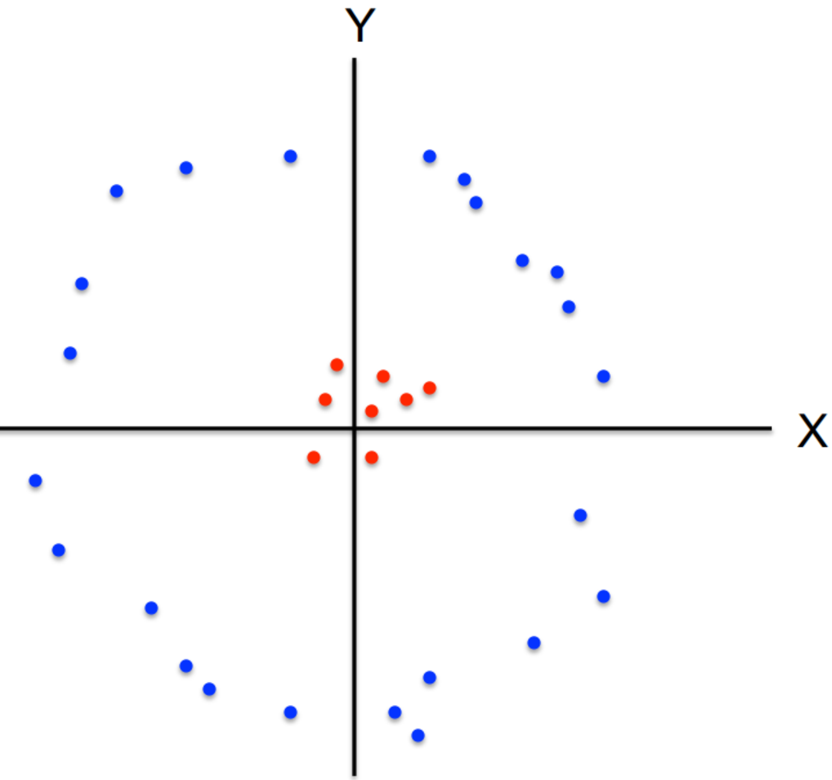
Would be better to have
one cluster here

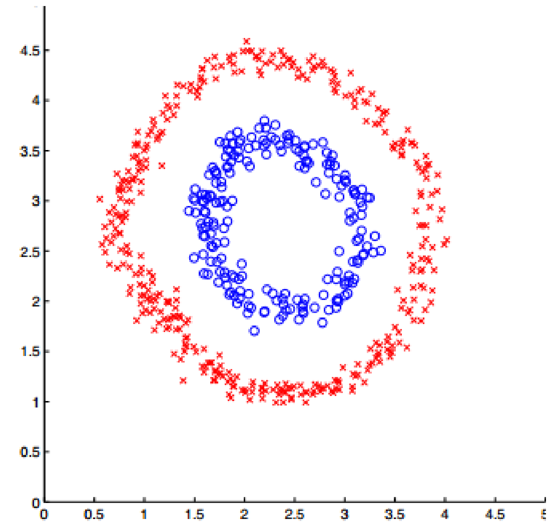


… and two clusters here

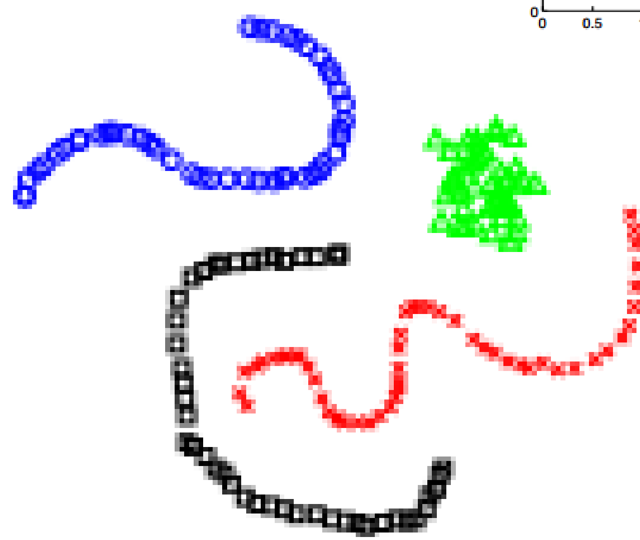
We still have ***problem*** *of* ***local minima.***

Spectral Clustering





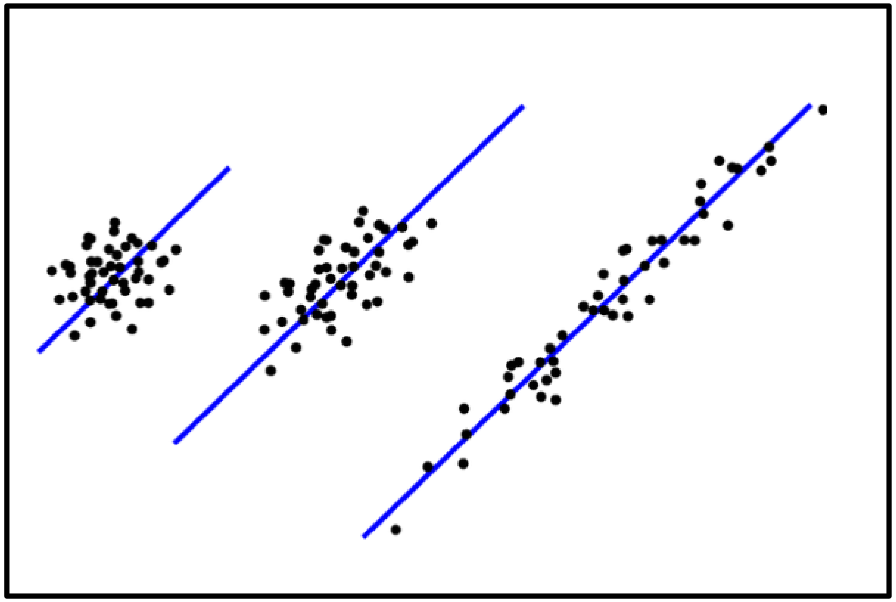
(e)

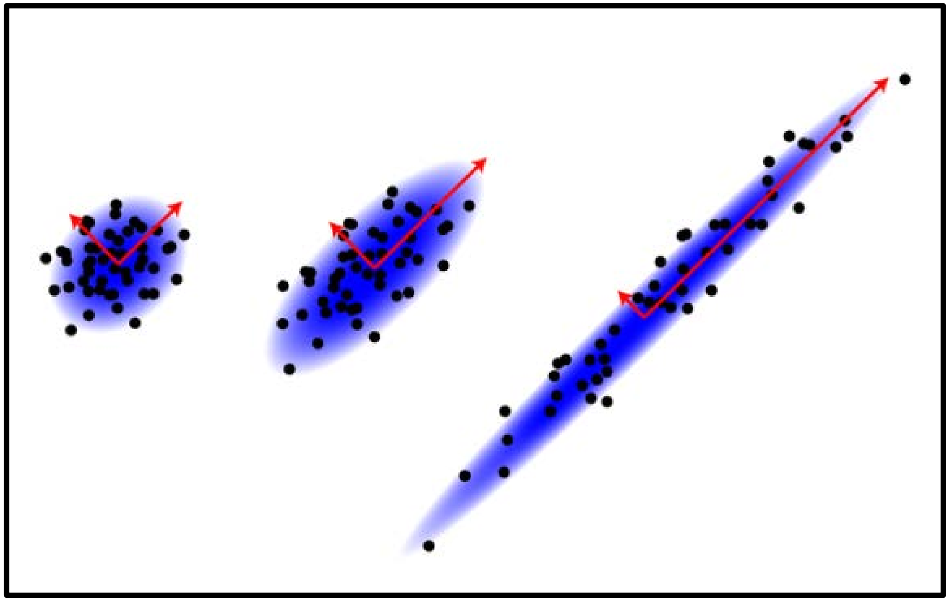


Part VI – Principal Component Analysis

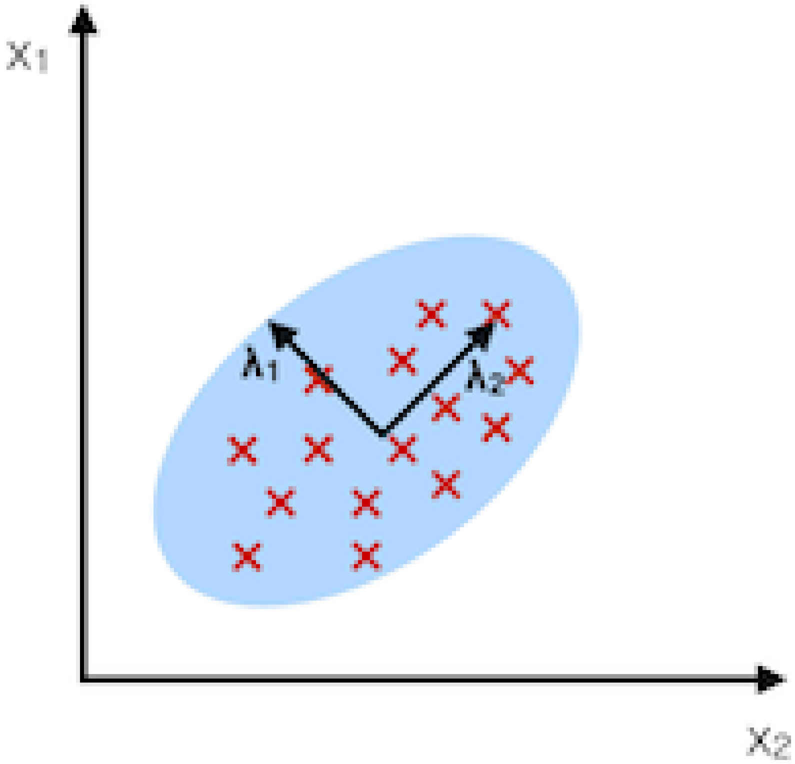
Principal Component Analysis

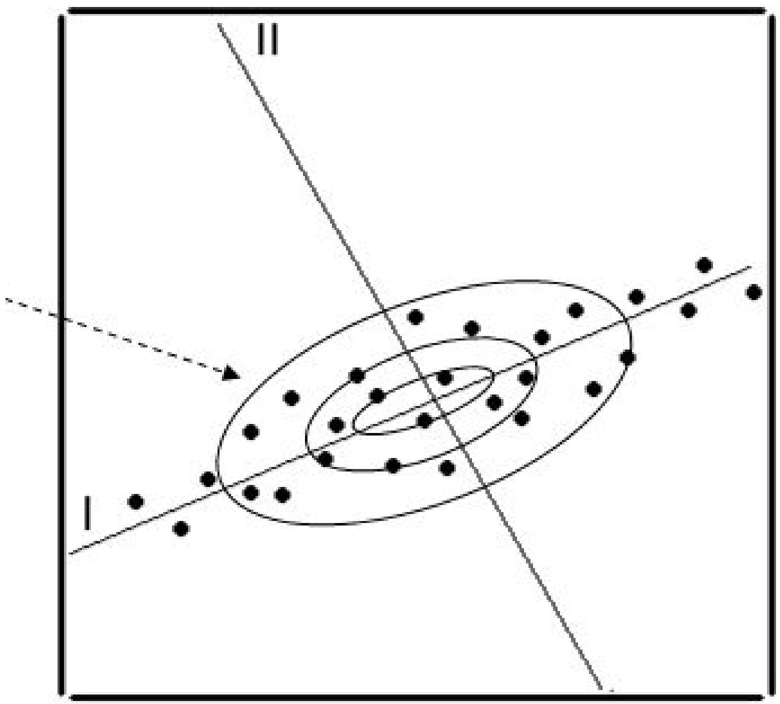
- PCA reduces the **dimensionality of a dataset**.
- **Dimensionality** = the number of variables
- Two variables **principal components**.
- The **first component:** represents the direction of the highest variance of the data.
- The **second component**: represents the highest of the remaining variance.





Principal Component Analysis

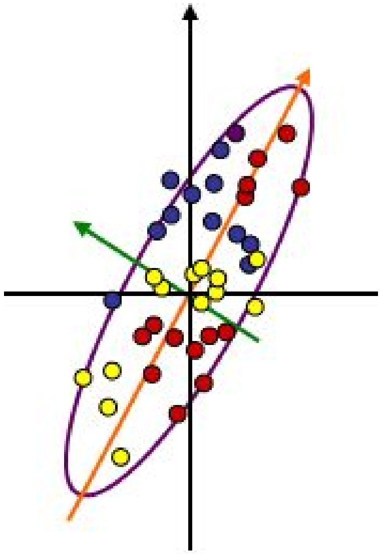




- **Component axes** that maximize the variance
- **Eigenvalue** tells how much the **variance** is.
- **Eigenvector** tells the **direction** of the variation

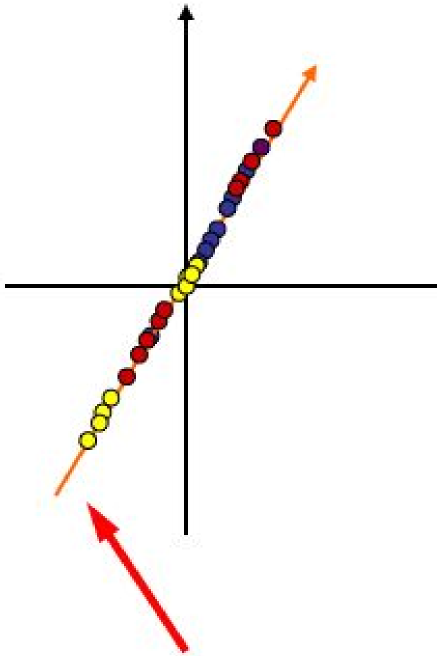
Principal Components Analysis

- Searching for **spatial directions** having **highest variance.**
- **Project the data** onto the subspace of highest variance
- Structure encoded in the **sample co-variance of the data.**



$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

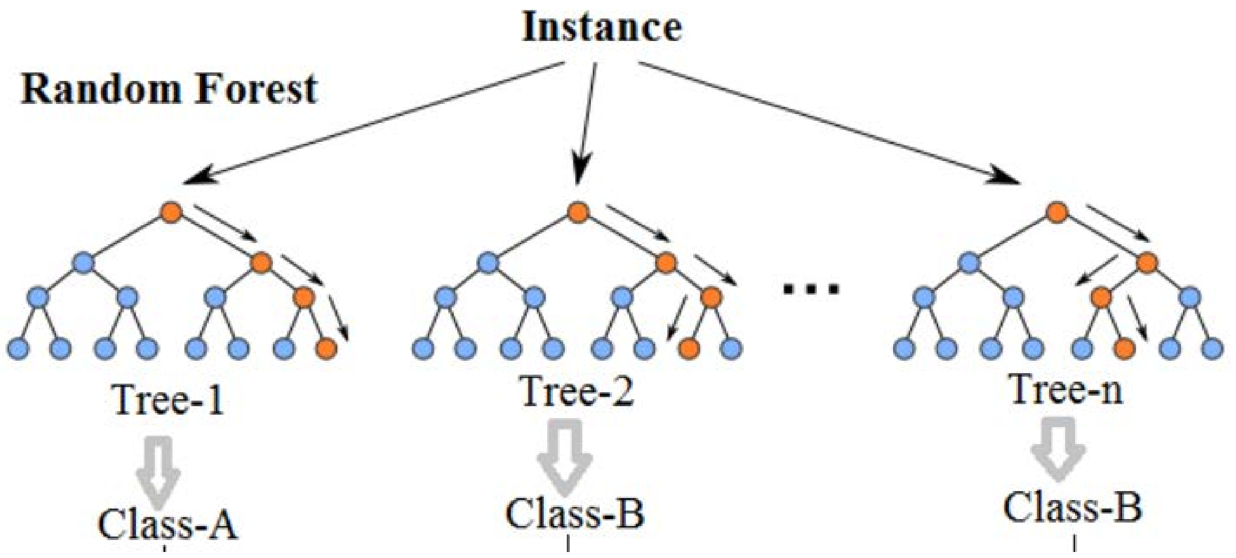
$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T$$



Part VII – Random Forest

Random Forest

- Random Forest is a statistical algorithm that is used to cluster.
- Random forests are a combination of **tree predictors.**
- Each tree depends on the values of a **random vector.**
- The tree with the **most predictive power** is shown as output.
- Forest: The program makes **multiple trees**
- Each **tree is different** because for each split in a tree, variables are chosen at random.



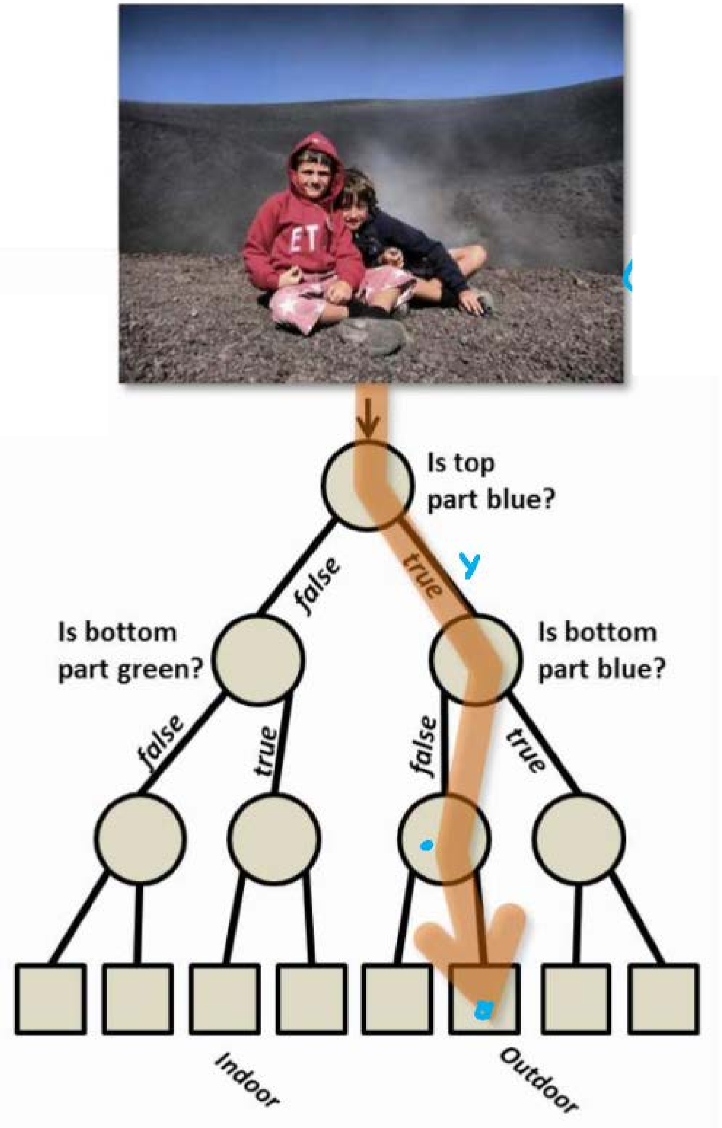
Decision Tree

- When the data set is large and/or there are many variables it becomes difficult to cluster the data.

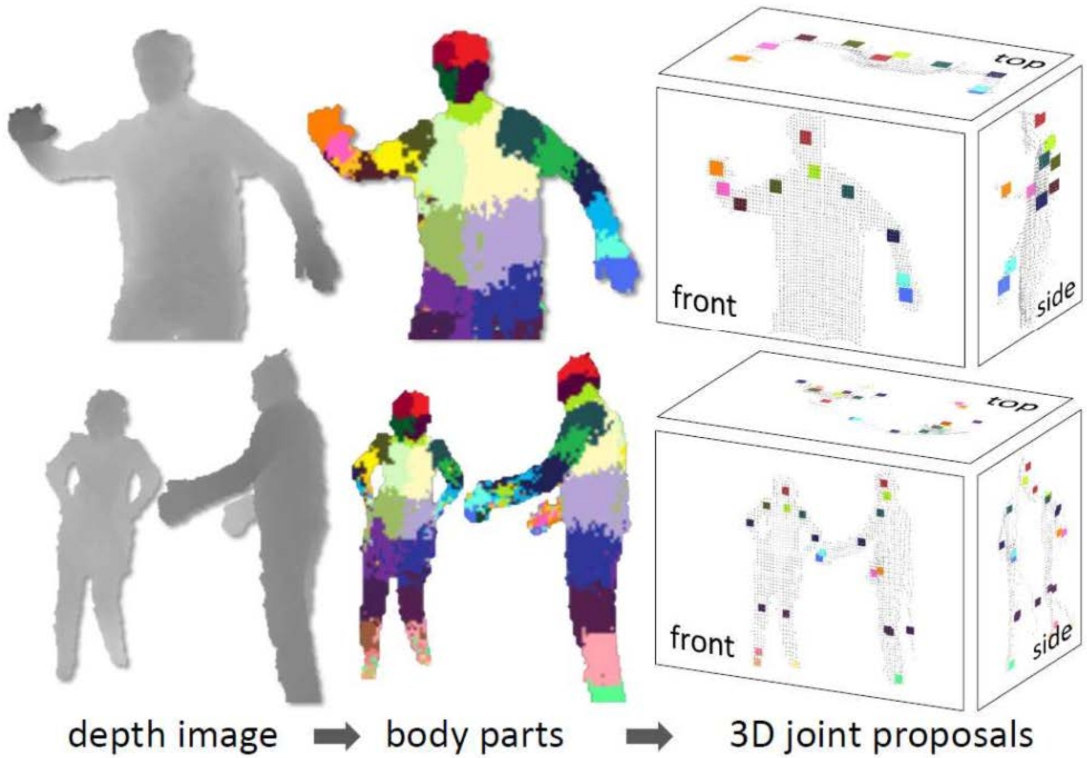
- **Random forests** is **non-parametric** because it is not based on any assumptions about data distribution.

- The algorithm **clusters the data in groups** and subgroups.

- The structure look like a tree. This is called a decision tree.



Random Forests and Kinect



[Jamie Shotton et al 2011]

Dimensionality reduction = generate a large and carefully constructed **set of trees.**

Lab Activities

- Activity 1: K-Means (45min)
- Activity 2: K Nearest Neighbours (45min)

- Break

- Activity 3: Writing (30min)
- Extra Activity 4: Principal Component Analysis (30min)

References

[1] - Machine Learning: Decision Trees - Chuck Dyer
[2] - Introduction to Learning and Decision Trees - Sandholm and Sing Lee
[3] - Machine Learning - Milos Hauskrecht
[4] - Supervised learning - Andrew Ng
[5] - A Gentle Introduction to Support Vector Machines in Biomedicine - Statnikov et al
[6] - Mathematics for Inference and Machine Learning - Deisenroth and Zafeiriou
[7] - Support Vector Machine (and Statistical Learning Theory) - Jason Weston
[8] - Weighted K Nearest Neighbor - Siddharth Deokar
[9] - Seven Techniques for Dimensionality Reduction - Silipo et al
[10] - Machine Learning - Nando de Freitas
[11] - Machine learning - Kevin Murphy
[12] - Machine Learning - Tommi Jaakkola
[13] - Machine Learning - A. Zisserman
[14] - The SVM classifier - A. Zisserman
[15] - Introduction to Learning - David Sontag
[16] - Statistical and Machine Learning - Su-Yun Huang
[17] - Lecture Notes in Machine Learning - Zdravko Markov
[18] - Semi-Supervised Learning - Chapelle et al
[19] - Learning theory - David Sontag
[20] - Introduction to Machine Learning - Miguel A. Carreira-Perpinan

References

[21] - Machine Learning and Decision Trees - Lars Schmidt-Thieme

[22] - An introduction to random forests - Eric Debreuve

[23] - Random Forestes - Leo Breiman

[24] - Support Vector Machines - Arthur Gretton

[25] - Support vector machines - Wisniewski and Wawrzyniak

[26] - Introduction to Machine Learning - Mehryar Mohri

[27] - Introduction, Regression Analysis, and Gradient Descent - Andrew Ng

[28] - Introduction to Machine Learning - Smola and Vishwanathan

[29] - Machine Learning - Random Decision Forests and Deep Neural Networks - Kari Pulli

[30] - Tutorial To Implement k-Nearest Neighbors in Python From Scratch - Jason Brownlee

[31] - Machine Learning. K-means, kNN - KTH Royal Institute Of Technology

[32] - Introduction to Machine Learning - Nils J. Nilsson

[33] - Induction of Decision Trees - J.R. Quinlan

[34] - CS 446: Machine Learning - Dan Roth - University of Illinois, Urbana-Champaign

[35] - PCA - Principal Component Analysis - Matthias Scholz, Ph.D. thesis

[36] - Machine Learning / Course Overview & Lecture 1 - Agarwal and Bhattacharya

[37] - Unsupervised learning: Clustering, Ata Kaban, The University of Birmingham

[38] - CS229 Lecture notes - Andrew Ng

[39] - The Expectation-Maximisation (EM) algorithm, Alastair Turl, School of Computer Science

[40] - K-Nearest Neighbours and Instance based learning , Ata Kaban

References

[41] - Machine Learning / Convnets - Nando de Freitas

[42] - Bayesian Machine Learning - Lecture 1 - Guido Sanguinetti

[43] - Introduction to Machine Learning - Lecture 1 : Overview - Chaohui Wang

[44] - Machine Learning (Extended) - Dr. Ata Kaban

[45] - CS 540 Lecture Notes - C. R. Dyer - Machine Learning (Chapter 18.1 - 18.3)

[46] - The Expectation-Maximisation (EM) algorithm, Alastair Turl

[47] - Random Forest Algorithm, An Interactive Discussion, Niraj Kumar

[48] - Random Forests - Grant Hamilton

[49] - shapeofdata / Neuron system

[50] - Linear Separation and Support Vector Machines - Jesse Johnson

[51] - Visualization of machine-learning classification trees, Tae-Kyun Kim and Bjorn Stenger

[52] - Principal Component Analysis, Jesse Johnson

[53] – Stanford Machine Learning