**OVERVIEW OF DATAMINING**

# Pattern Discovery techniques

Pattern discovery is one of the primary steps in data mining. There are different types of learning techniques applied in pattern discovery.

1. Supervised Learning
2. Unsupervised Learning
3. Semi-supervised Learning
4. Active Learning
5. Online Learning
6. Deep Learning

## 1. Supervised Learning

* Supervised learning deals with labeled data with training examples
* Main problems that it solves:
  + Prediction
    - Algorithms used: Regression
  + Classification: Classification techniques can be grouped as below
    - Model-based learning: Where models are built early and applied on new cases
      * Algorithms used:
        + Decision trees
        + Linear Discriminant analysis
        + Support Vector machines
        + Neural networks
    - Instance based learning: Where no model is built in prior
      * K nearest neighbors
    - Probabilistic models:
      * Algorithms can be any of the above as long as the output is probabilistic rather than deterministic; in addition, Bayes classifier can be added to this category

## 2. Unsupervised Learning

* Deals with data which do not have labels
* Main problems that it solves:
  + Clustering (Algorithm: K-means)
  + Dimensionality reduction (Algorithm: Principal component analysis)
  + Association rule mining (Algorithm: A-priori)

## 3. Semi-supervised Learning

* Only a subset of data has labels and the remaining data is unlabeled labels
* Use algorithms from supervised learning to learn labels for unlabeled data that you feed in continuously to provide learning for the entire data

## 4. Active Learning

* Inclusion of a human expert to provide label
* Asking an expert to provide labels for millions of cases is not feasible. Hence a small set of representative examples from the data should be selected and provided to the experts to obtain labels for them
* The process of selecting the most representative cases in the data constitutes active learning

## 5. Online Learning

* Updating the learning model as more labels become available over time; it is especially useful for streaming data

## 6. Deep Learning

* Learning includes multiple layers and takes advantage of abundant cheap computational tools.
* Algorithms include neural network with multiple layers.
* Eliminates the need to extract features. Raw data can be used.

# Challenges in Classification

There are several challenges when performing classification techniques.

## Hard to get high performance using a single classifier

A classification problem yields different accuracies when using different techniques.

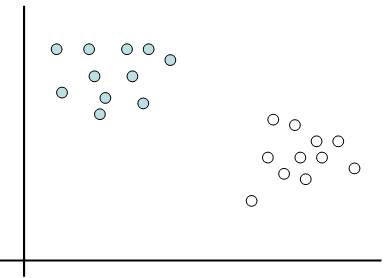
**Solution: Ensemble of classifiers**

Single classifiers are weak. But when combined with other classifiers they become provide high performance

## Not all the problems deal with linearly separable data

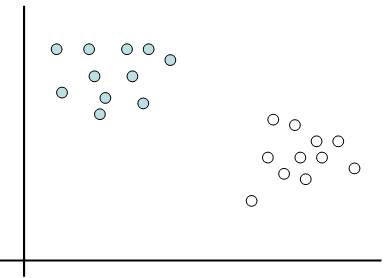
Decision tree

Let us consider the below distribution of data. In this case the data points can be clearly separated using Feature F1 (Orange line). A decision tree will use IF-THEN rule based on F1 value to classify the data points.

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiK9d6RipjKAhVBSCYKHQtUDCMQjRwIBw&url=https://onionesquereality.wordpress.com/2009/03/22/why-are-support-vectors-machines-called-so/&psig=AFQjCNHwK6ttOaYsUysyHs-sIq5kuZQebQ&ust=1452267683897456)F1

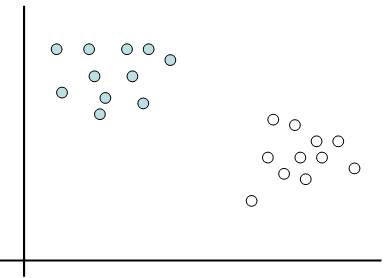
LDA

Using LDA, the data points can be separated as shown in orange line

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiK9d6RipjKAhVBSCYKHQtUDCMQjRwIBw&url=https://onionesquereality.wordpress.com/2009/03/22/why-are-support-vectors-machines-called-so/&psig=AFQjCNHwK6ttOaYsUysyHs-sIq5kuZQebQ&ust=1452267683897456)F1

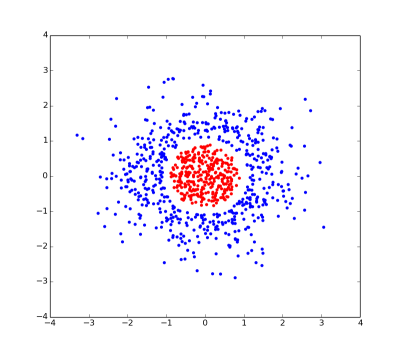
SVM

Using SVM the data points can be separated as shown in orange line (Hyperplane).

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiK9d6RipjKAhVBSCYKHQtUDCMQjRwIBw&url=https://onionesquereality.wordpress.com/2009/03/22/why-are-support-vectors-machines-called-so/&psig=AFQjCNHwK6ttOaYsUysyHs-sIq5kuZQebQ&ust=1452267683897456)F1

Since the above data points were linearly separable, decision tree, LDA, SVM provided good results.

LINEARLY INSEPARABLE DATA

[](https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwjoptOLj5jKAhXLKCYKHWqfBosQjRwIBw&url=http://stackoverflow.com/questions/1148513/difference-between-a-linear-problem-and-a-non-linear-problem-essence-of-dot-pro&psig=AFQjCNHwK6ttOaYsUysyHs-sIq5kuZQebQ&ust=1452267683897456)

**Solution: Linearly inseparable data can be transformed to be linearly separable using Kernel Models.**

## The data is not always deterministic (Uncertainty in the data)

The outcome of the classifier need not always be deterministic. Some applications may require models to give them the confidence percentage for a case belonging to a particular class.

**Solution: Probabilistic Classifiers (Bayes classifier)**

In addition to learning, the relationship between data has to be represented. This can be done with **probabilistic graphical models** like

* Bayesian networks (If the direction of relationship is important)
* Markov models (If direction of relationship is not required)

## Time

Most of the models do not include time component. If the time component is included, then **dynamic probabilistic graphical models (Dynamic PGM)**