Introduction to Pattern Recognition and Machine Learning

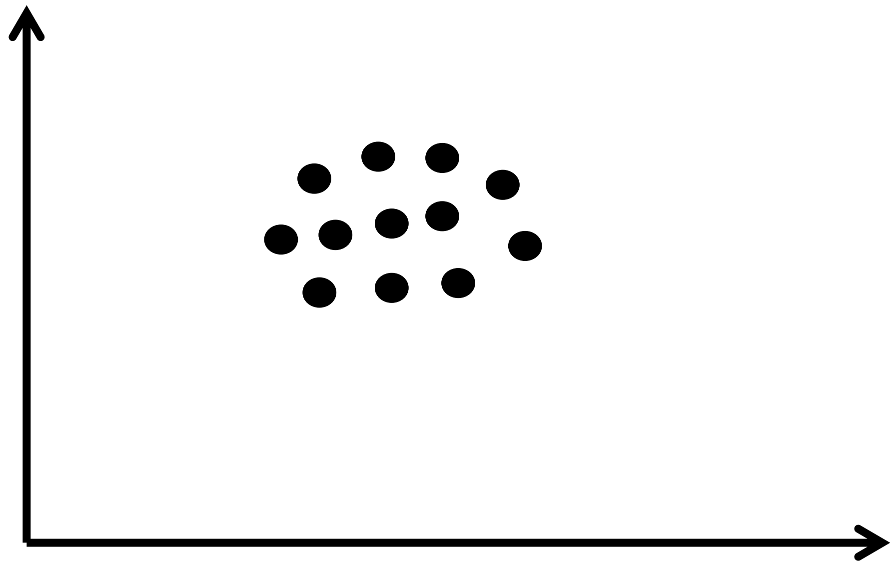
Alexandros Iosifidis
Academy of Finland Postdoctoral Research Fellow
(term 2016-2019)

TAMPERE
UNIVERSITY OF
TECHNOLOGY

Tampere, August 2016

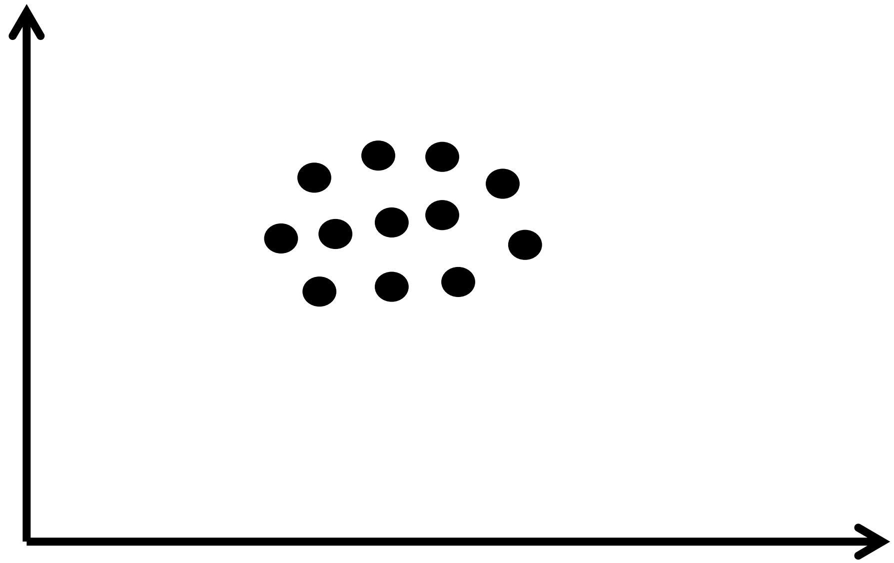
Vector Spaces

- What do we observe here?



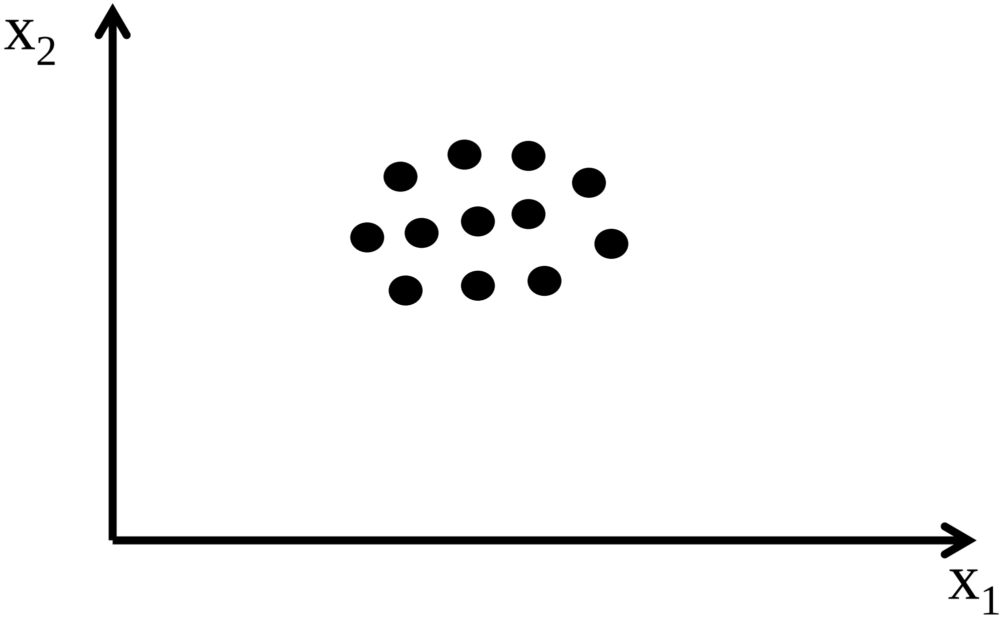
Vector Spaces

- What do we observe here?
  - A set of N points in the 2-dimensional space



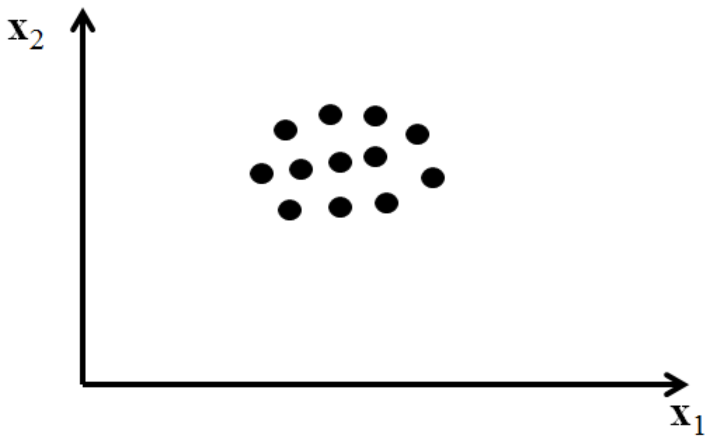
Vector Spaces

- What do we observe here?
  - A set of N points in the 2-dimensional space
  - Each point can be represented by a vector $\mathbf{x}_i=\{\mathbf{x}_{i1},\mathbf{x}_{i2}\}$, $i=1\ldots,N$



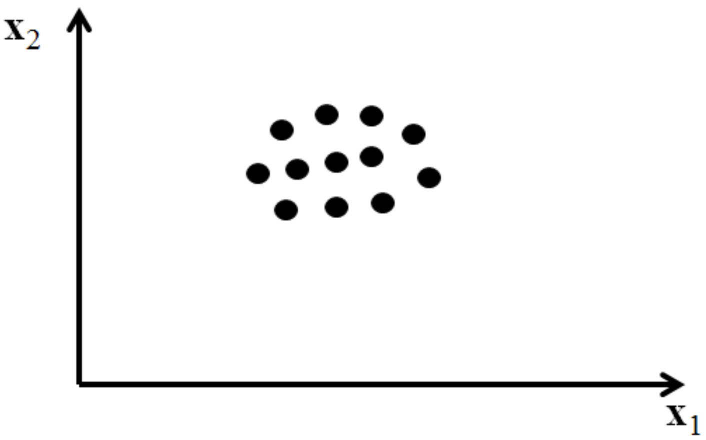
Vector Spaces

- What do we observe here?
  - A set of N points in the 2-dimensional space
  - Each point can be represented by a vector $\mathbf{x}_i=\{\mathbf{x}_{i1},\mathbf{x}_{i2}\}$, i=1…,N
  - We call each point as sample and we say that $\mathbf{x}_i$ is the representation of the i-th sample



Vector Spaces

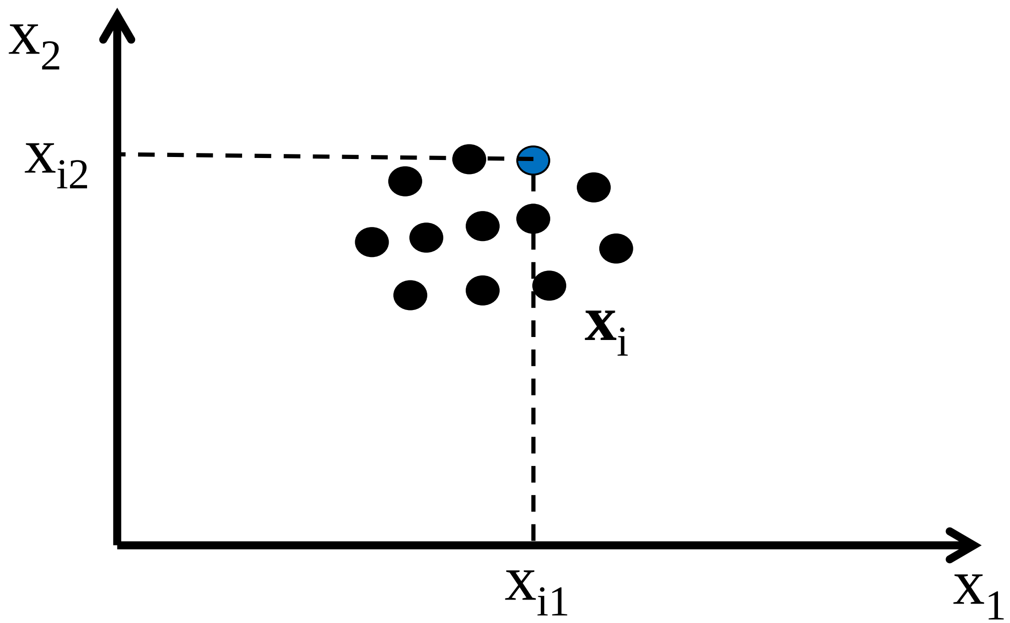
- What do we observe here?
  - A set of N points in the 2-dimensional space
  - Each point can be represented by a vector $\mathbf{x}_i=\{\mathbf{x}_{i1},\mathbf{x}_{i2}\}$, $i=1\ldots,N$
  - We call each point as sample and we say that $\mathbf{x}_i$ is the representation of the i-th sample
  - We say that $\mathbf{x}_i$ belongs to the 2D vector space or $\mathbf{x}_i \in \mathbb{R}^2$ or $\mathbf{x}_i \in \mathbb{R}^D$, D=2.



Vector Spaces

- What is a sample representation $\mathbf{x}_i$?
  - Data representation is followed by a description, e.g. we can describe a person by using his/hers height and weight. In that case:

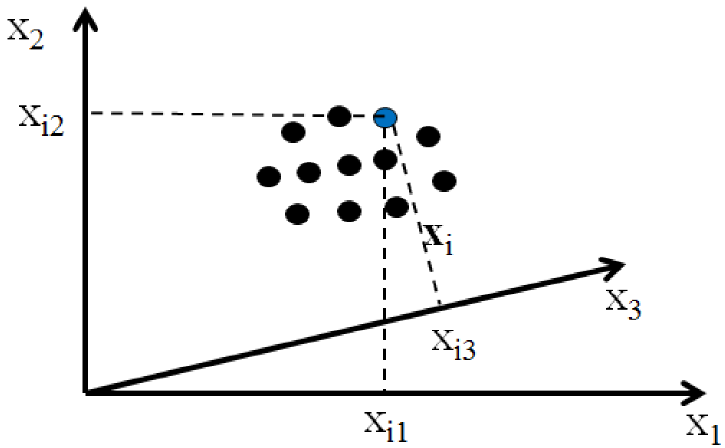
    Alekos → {187, 110}



Vector Spaces

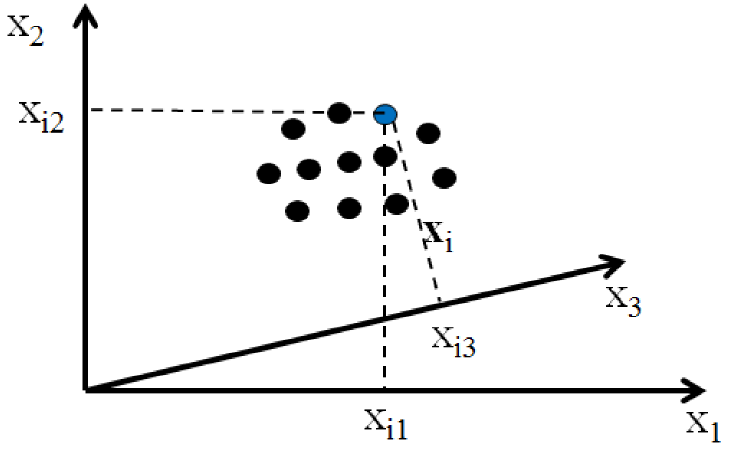
- What is a sample representation $\mathbf{x}_i$?
  - Another representation could be followed by the description of {height, weight, gender}. In that case: Alekos → {187, 110, 1}

Real value



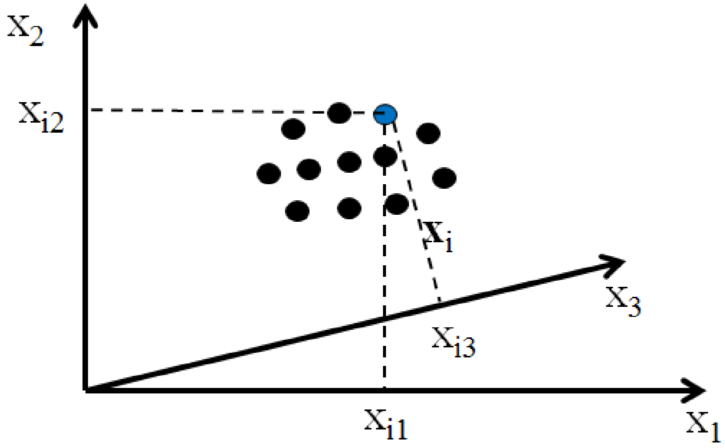
Vector Spaces

- What is a sample representation $\mathbf{x}_i$?
  - Another representation could be followed by the description of {height, weight, gender}. In that case: Alekos → {187, 110, 1}
  - Let us assume that all vectors represent males. Does it make sense to keep the last dimension?



Vector Spaces

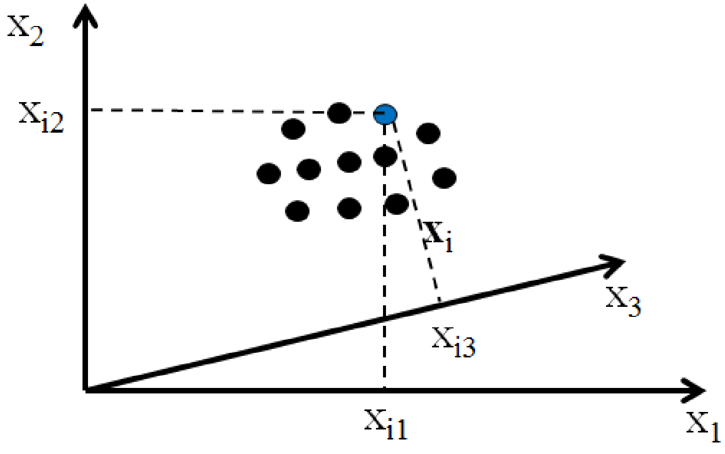
- What is a sample representation $\mathbf{x}_i$?
  - Another representation could be followed by the description of {height, weight, gender}. In that case: Alekos → {187, 110, 1}
  - Let us assume that all vectors represent males. Does it make sense to keep the last dimension?
  - Sometimes data dimensions contain low information.
  - Sometimes data dimensions can be combined in order to enhance properties of interest.



Vector Spaces

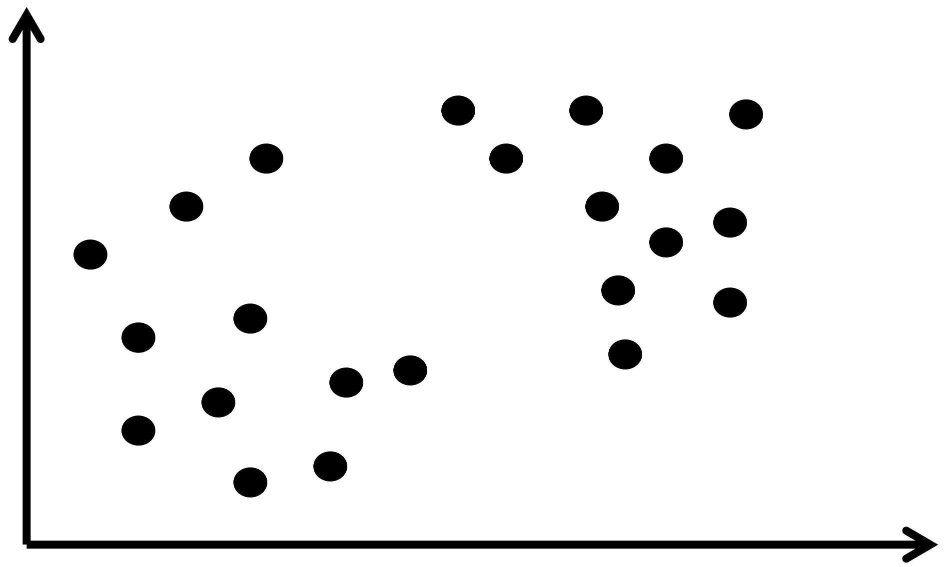
- What is a sample representation $\mathbf{x}_i$?
  - Another representation could be followed by the description of {height, weight, gender}. In that case: Alekos → {187, 110, 1}
  - Let us assume that all vectors represent males. Does it make sense to keep the last dimension?
  - Sometimes data dimensions contain low information.
  - Sometimes data dimensions can be combined in order to enhance properties of interest.

<u>Subspace learning</u>: Define a vector space $\mathbb{R}^d$, d < D, in which the data representations $\mathbf{y}_i \in \mathbb{R}^d$ have some properties of interest.

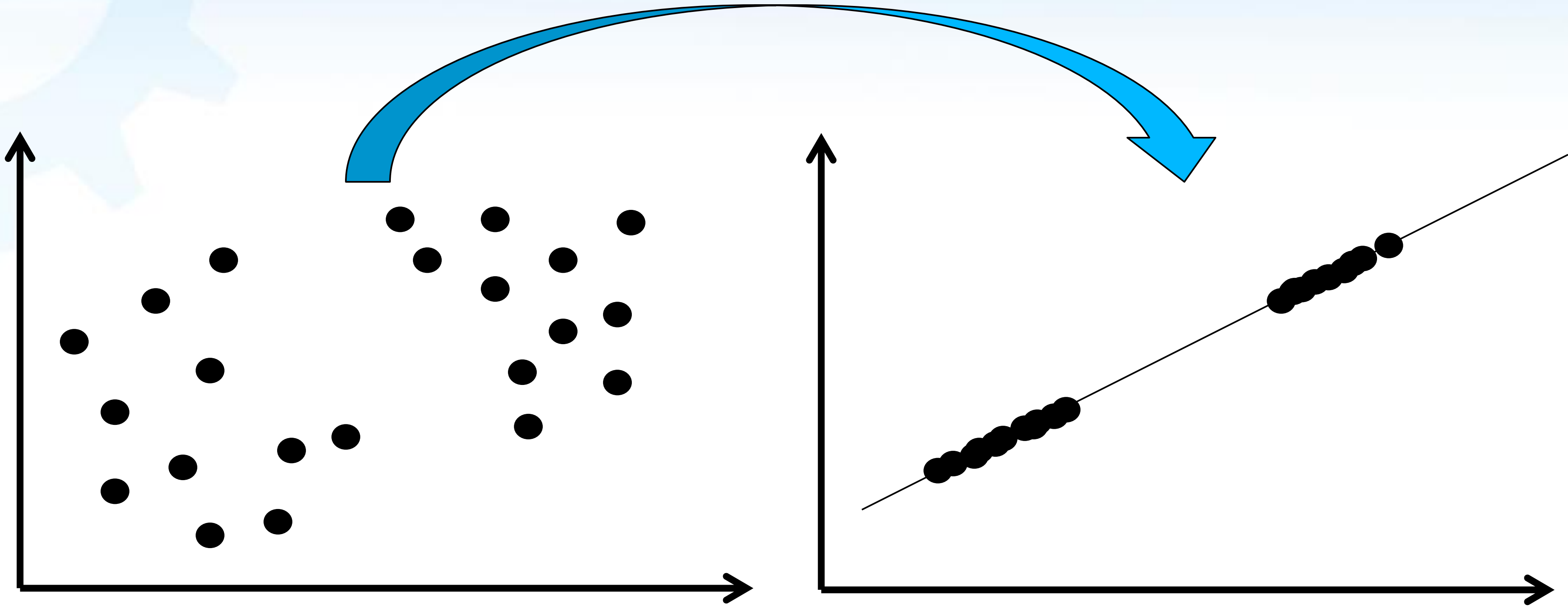


Vector Spaces

- Another example:



Vector Spaces

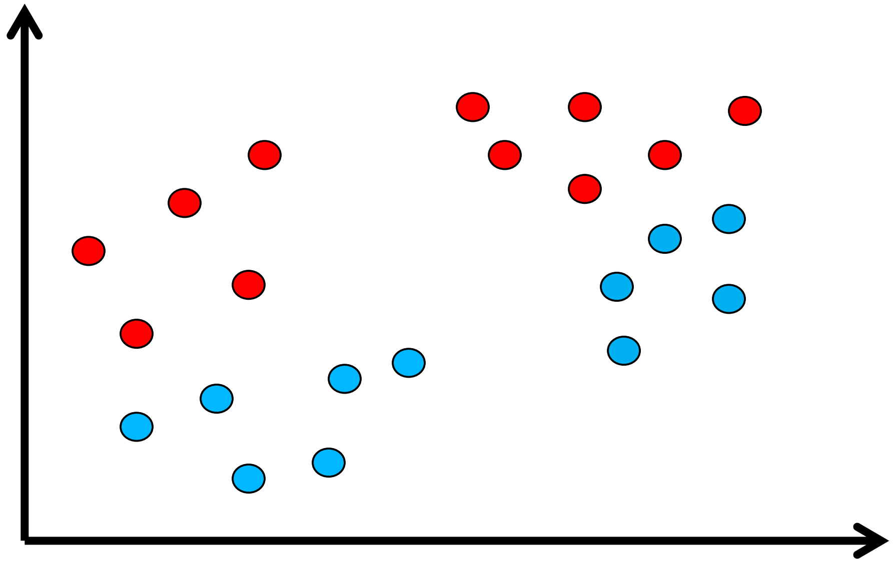


Input space R^D

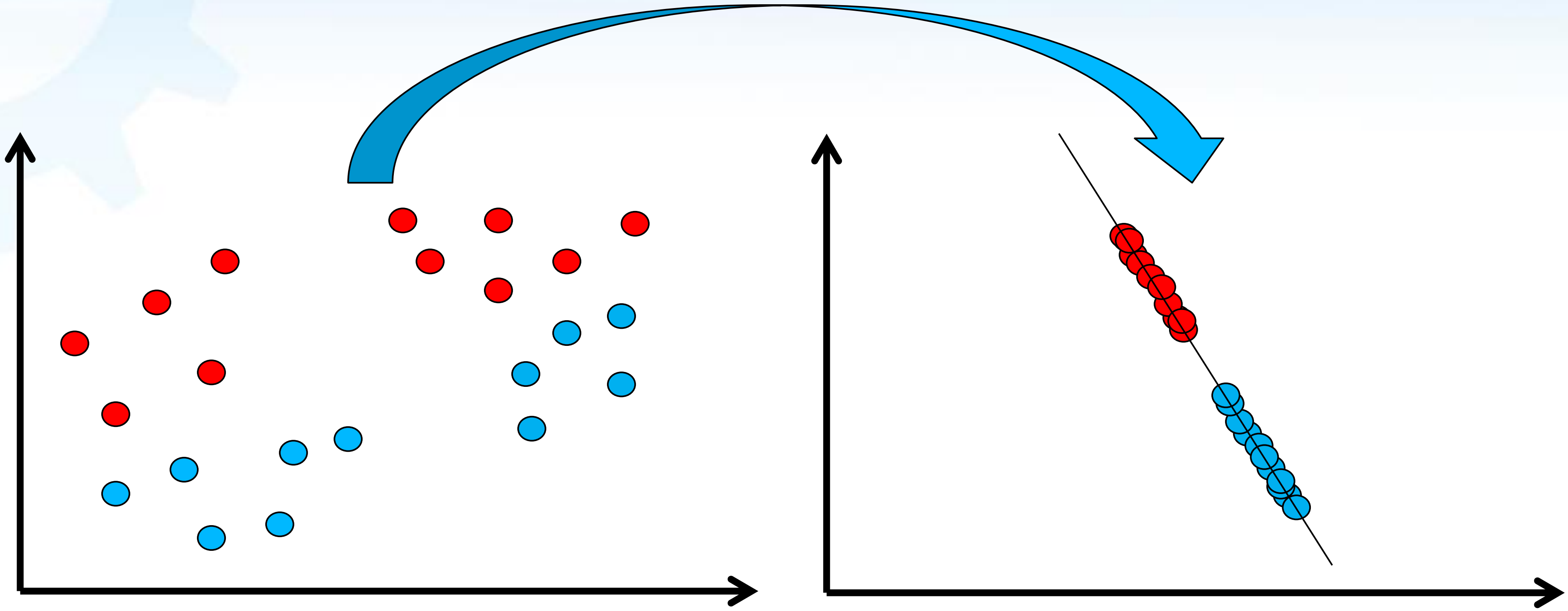
R^d, $d \leq D$ where data dispersion is maximized

Vector Spaces

- The same example:
  - Each sample is followed by a class label (red or green)



Vector Spaces



Input space R^D

R^d, $d \leq D$, where the dispersion within the classes is minimized and between the classes is maximized

Basic Problems in PR and ML

- Pattern Recognition problems can be divided in:
    - Unsupervised: the only available information is the data representations $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N

Basic Problems in PR and ML

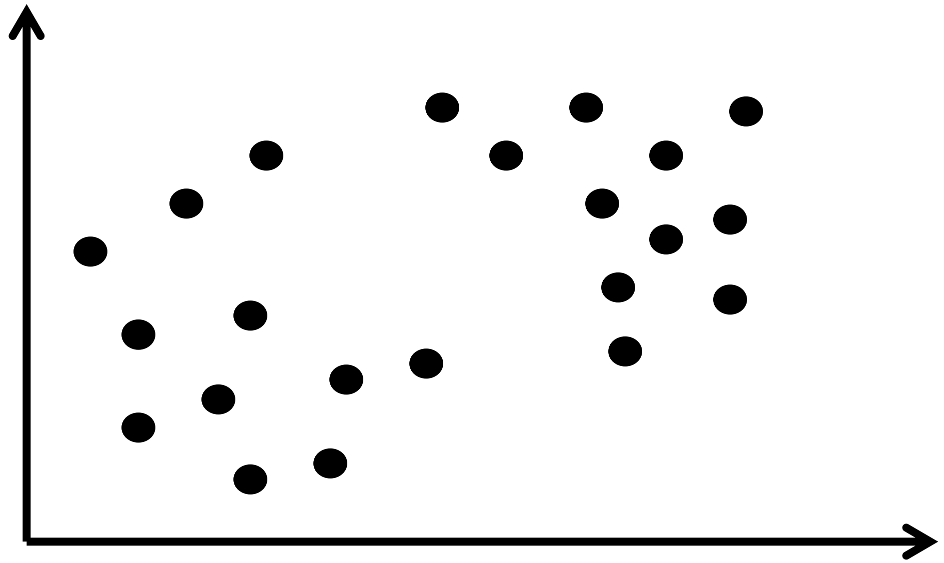
- Pattern Recognition problems can be divided in:
  - Unsupervised: the only available information is the data representations $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N
  - Supervised: data representations $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N and class labels l_i, i=1,…,N, $l_i \in \{1,\ldots,C\}$ are available

Basic Problems in PR and ML

- Pattern Recognition problems can be divided in:
  - Unsupervised: the only available information is the data representations $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N
  - Supervised: data representations $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N and class labels l_i, i=1,…,N, $l_i \in \{1,\ldots,C\}$ are available
  - Semi-supervised: a (small) sub-set of the data are labeled, i.e.: $\mathbf{x}_i \in \mathbb{R}^D$, i=1,…,N and l_i, i=1,…,n<N, $l_i \in \{1,\ldots,C\}$ are available

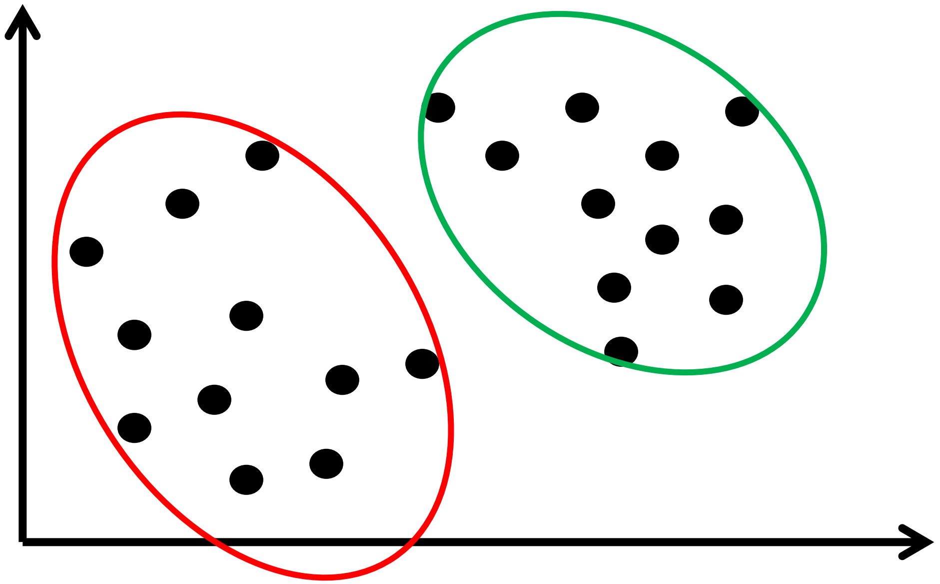
Basic Problems in PR and ML

- Data clustering:
  - Define a set of data groups (clusters) according to a (data closeness/similarity) criterion
  - Unsupervised problem



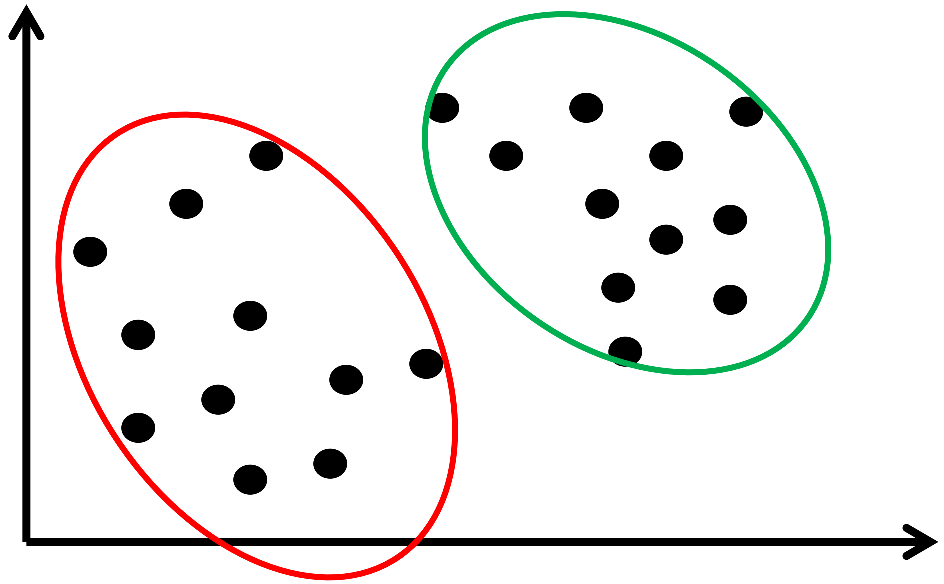
Basic Problems in PR and ML

- Data clustering:
  - Define a set of data groups (clusters) according to a (data closeness/similarity) criterion
  - Unsupervised problem



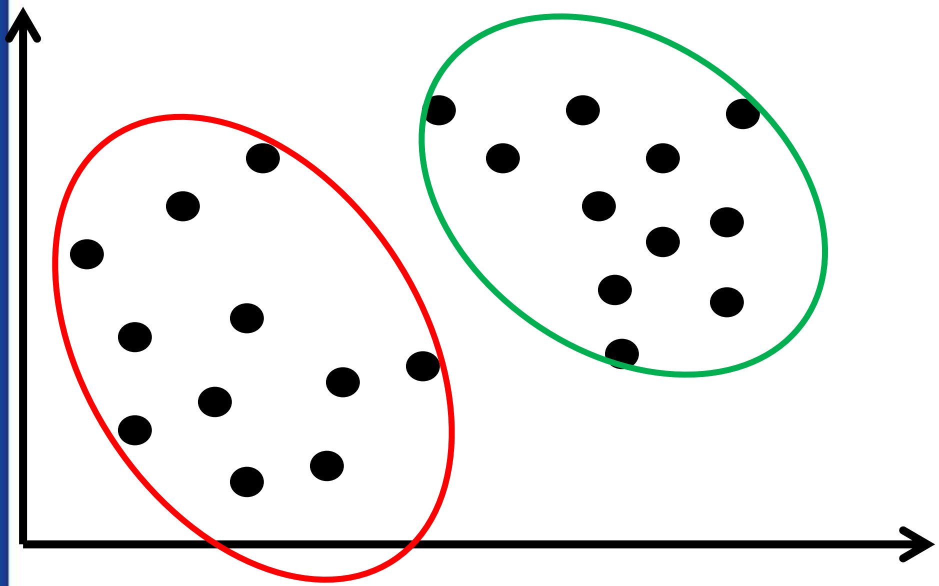
Basic Problems in PR and ML

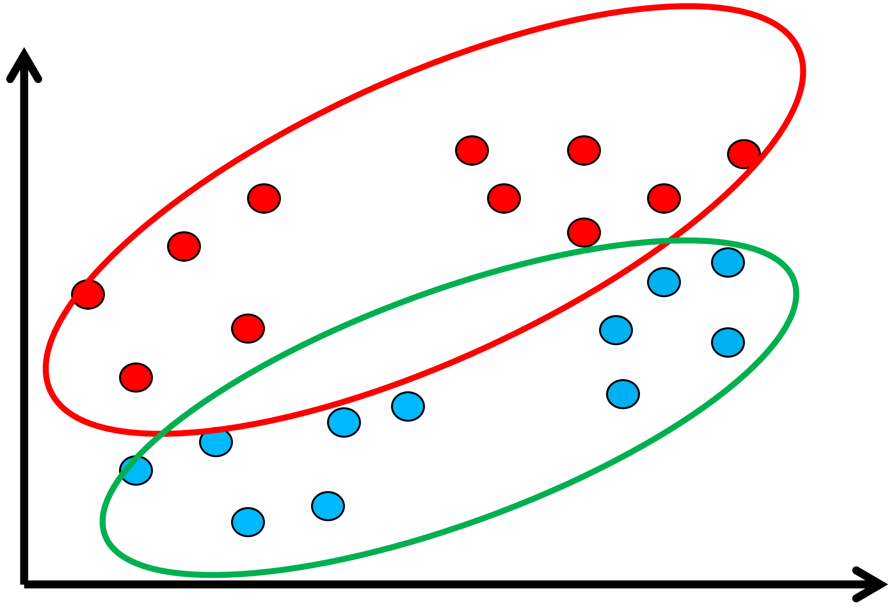
- Data clustering:
  - Define a set of data groups (clusters) according to a (data closeness/similarity) criterion
  - Unsupervised problem
  - Is this the correct answer?



Basic Problems in PR and ML

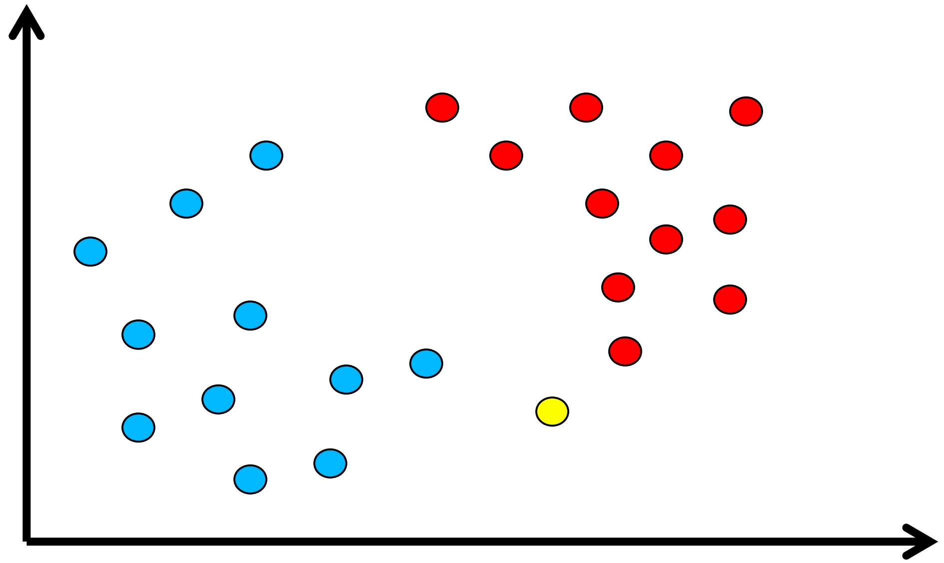
- Data clustering:
  - Define a set of data groups (clusters) according to a (data closeness/similarity) criterion
  - Unsupervised problem
  - Is this the correct answer?





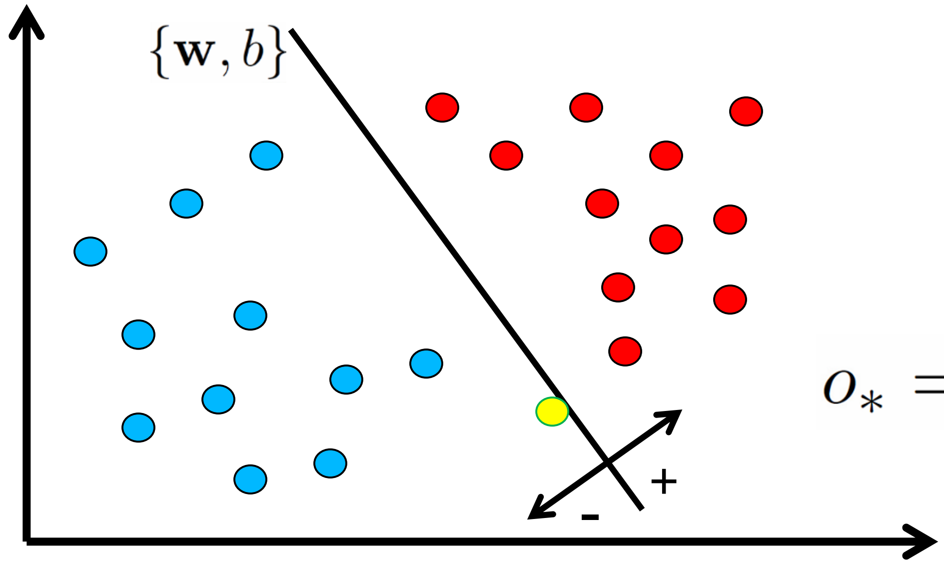
Basic Problems in PR and ML

- Data classification:
  - Samples belonging to two patterns (classes)
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



Basic Problems in PR and ML

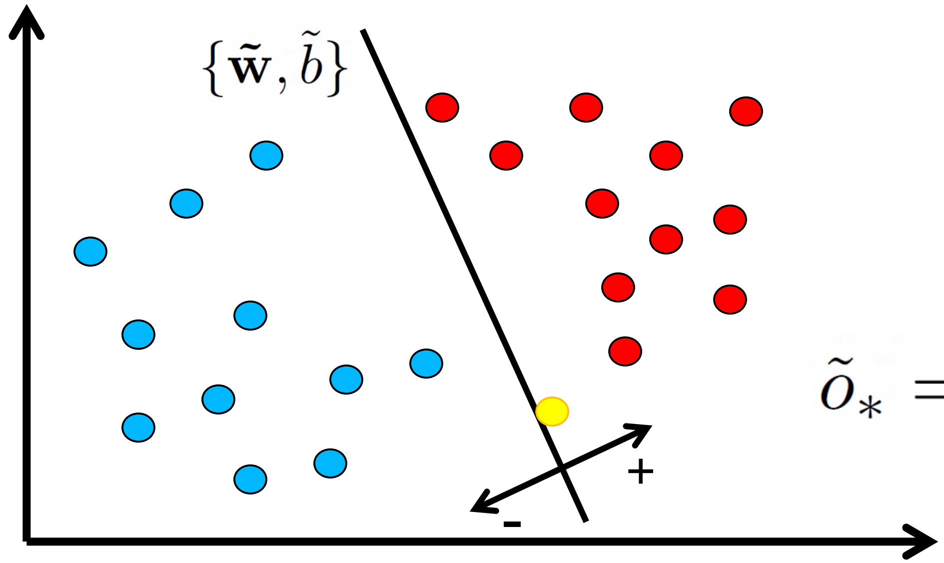
- Data classification:
  - Samples belonging to two patterns (classes)
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



$$o_* = \mathbf{w}^T \mathbf{x}_* + b$$

Basic Problems in PR and ML

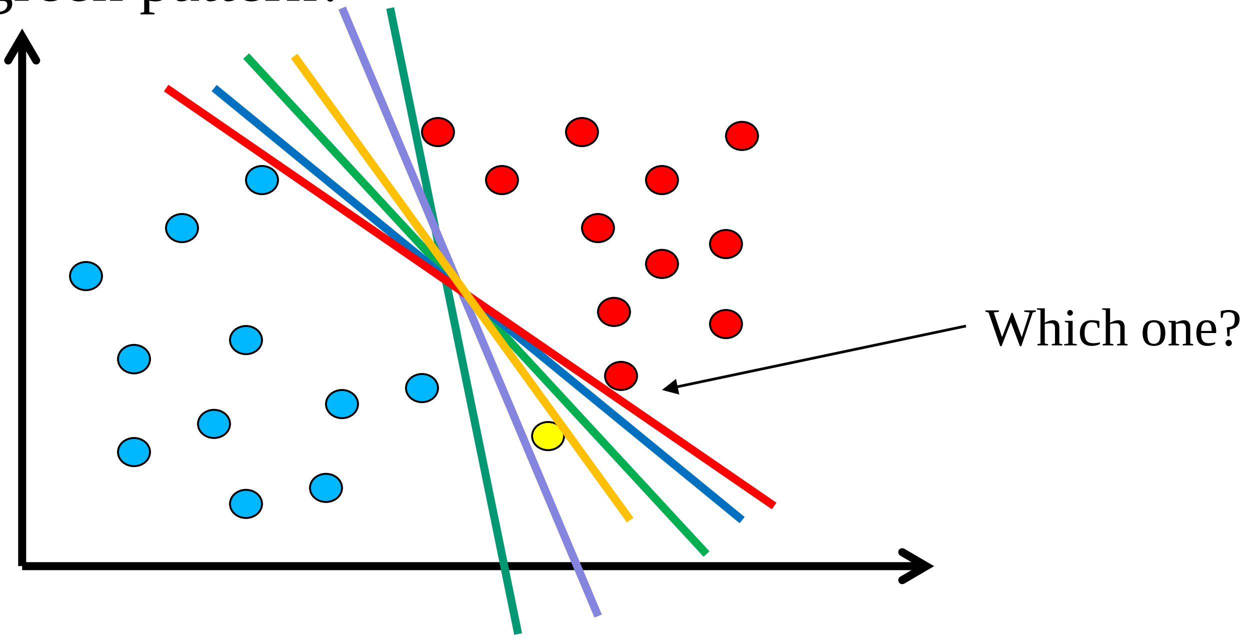
- Data classification:
    - Samples belonging to two patterns (classes)
    - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



$$\tilde{o}_* = \tilde{\mathbf{w}}^T \mathbf{x}_* + \tilde{b}$$

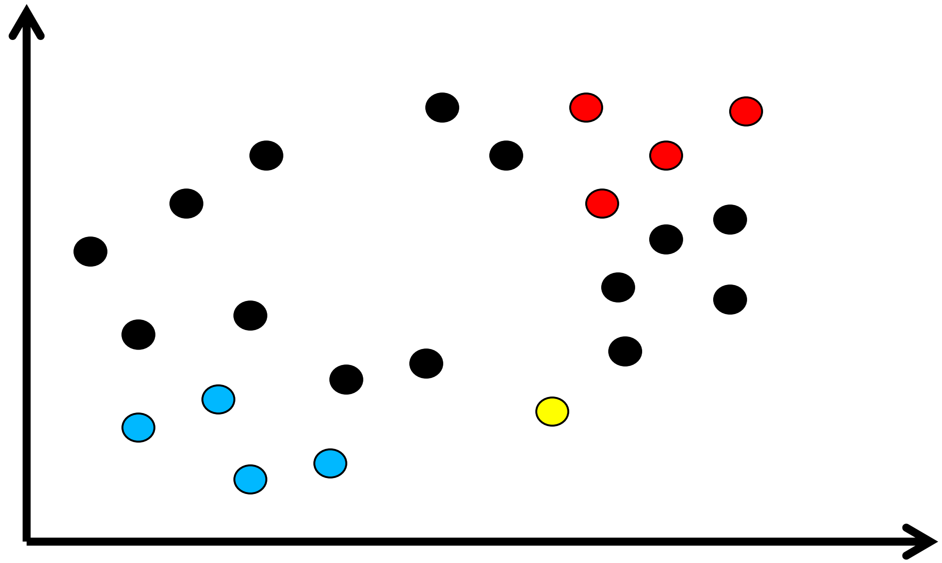
Basic Problems in PR and ML

- Data classification:
  - Samples belonging to two patterns (classes)
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



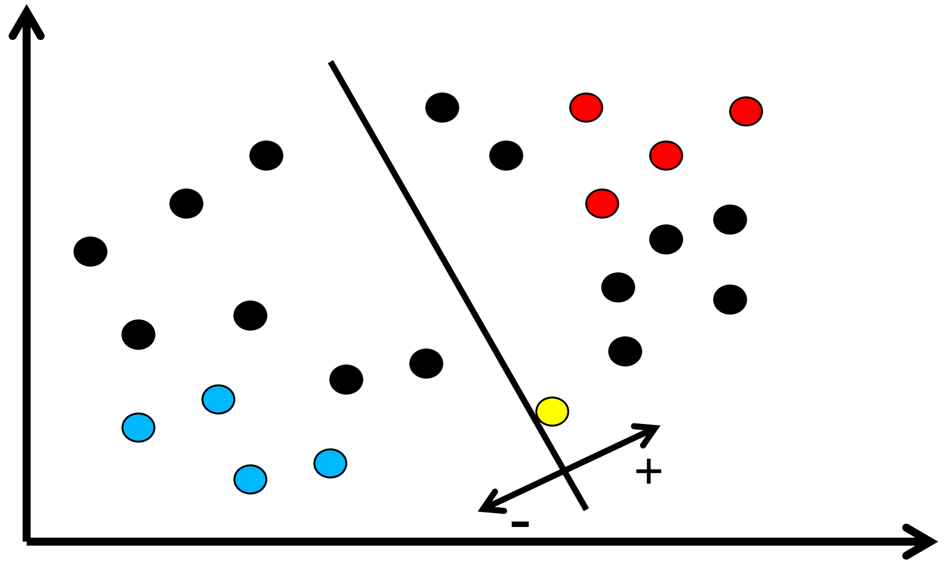
Basic Problems in PR and ML

- Data classification:
  - Some of the samples are followed by labels
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



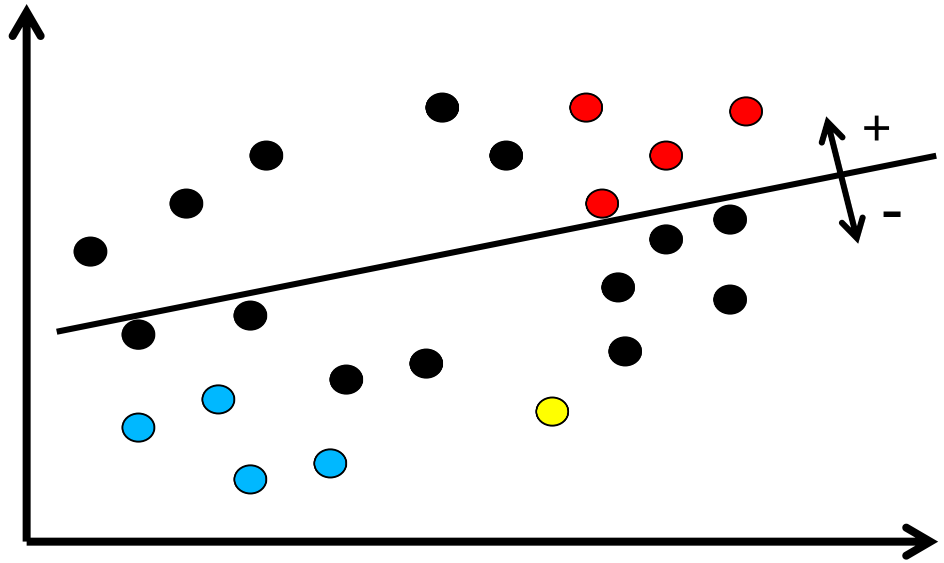
Basic Problems in PR and ML

- Data classification (semi-supervised):
  - Some of the samples are followed by labels
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



Basic Problems in PR and ML

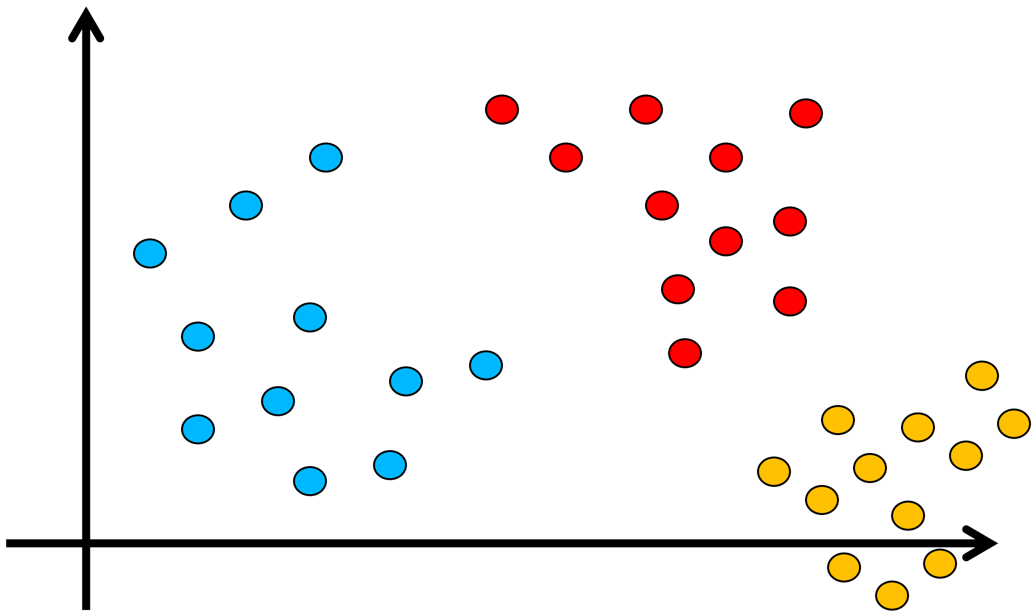
- Data classification (semi-supervised):
  - Some of the samples are followed by labels
  - Can we define a rule for recognizing if a new sample (the yellow one) belongs to the red or green pattern?



Many possible solutions!

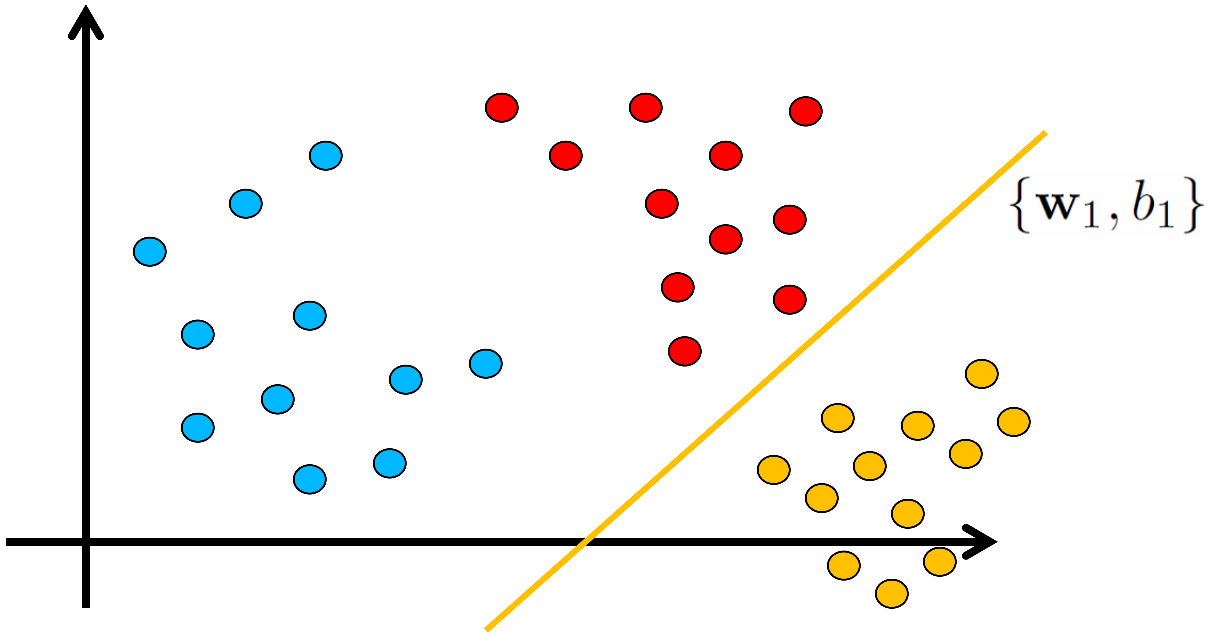
Nonlinearity

- Multi-class classification:
  - Split the problem in multiple binary (two-class) problems:
    - One-Versus-One and One-Versus-Rest



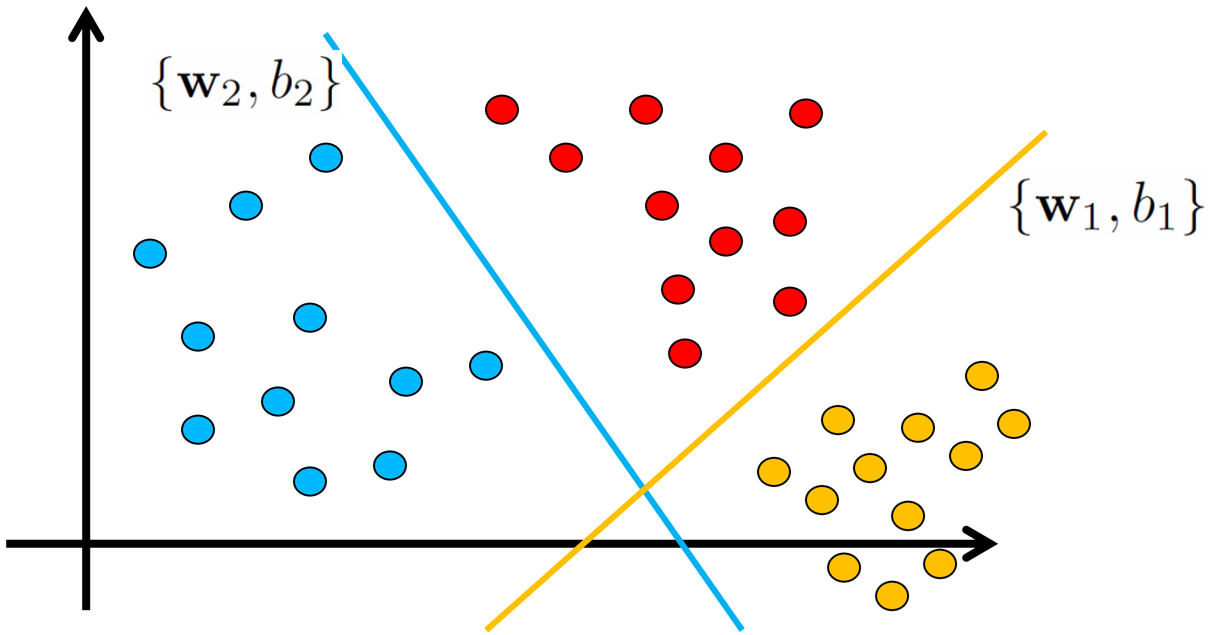
Nonlinearity

- Multi-class classification:
  - One-Versus-Rest case



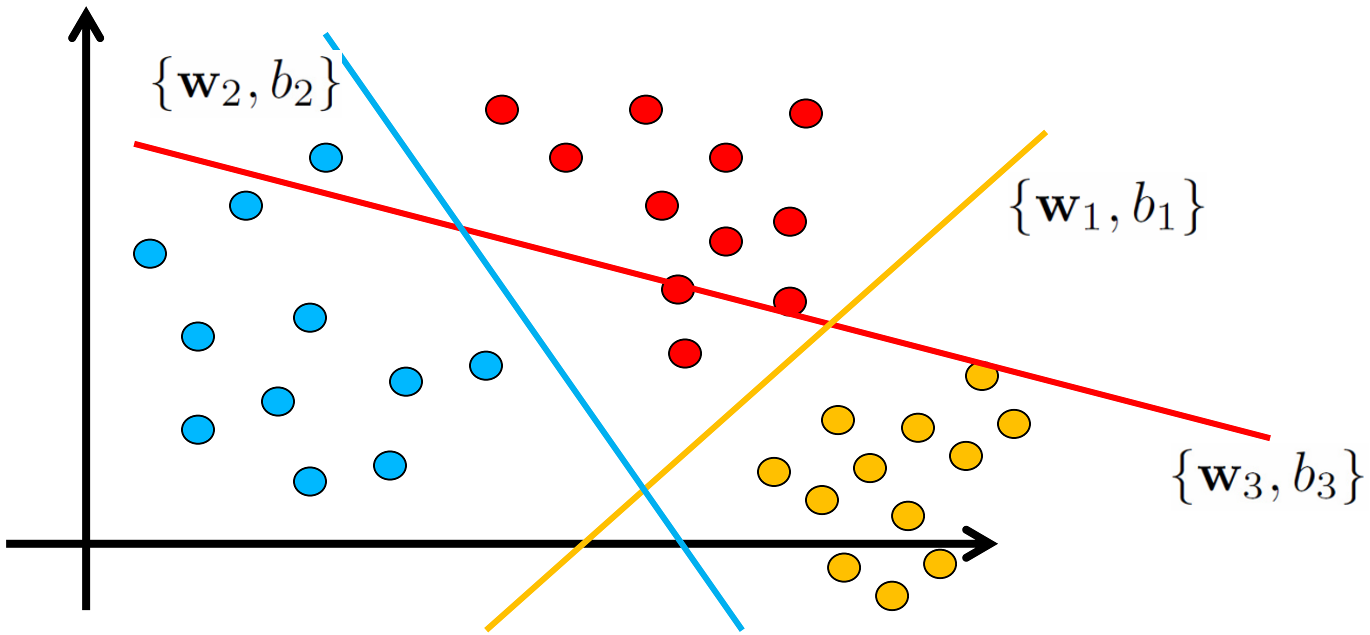
Nonlinearity

- Multi-class classification:
  - One-Versus-Rest case



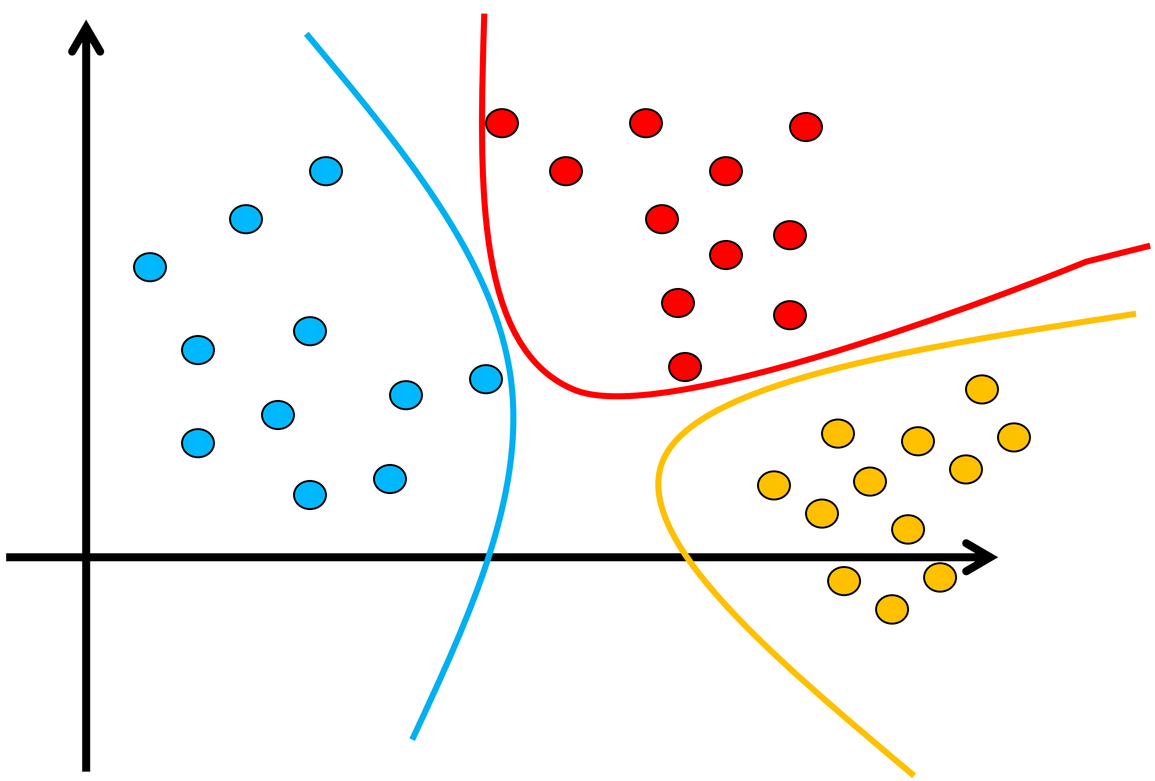
Nonlinearity

- Multi-class classification:
  - One-Versus-Rest case
  - Sometimes linear models are not the best solution



Nonlinearity

- Multi-class classification:
  - A better solution!

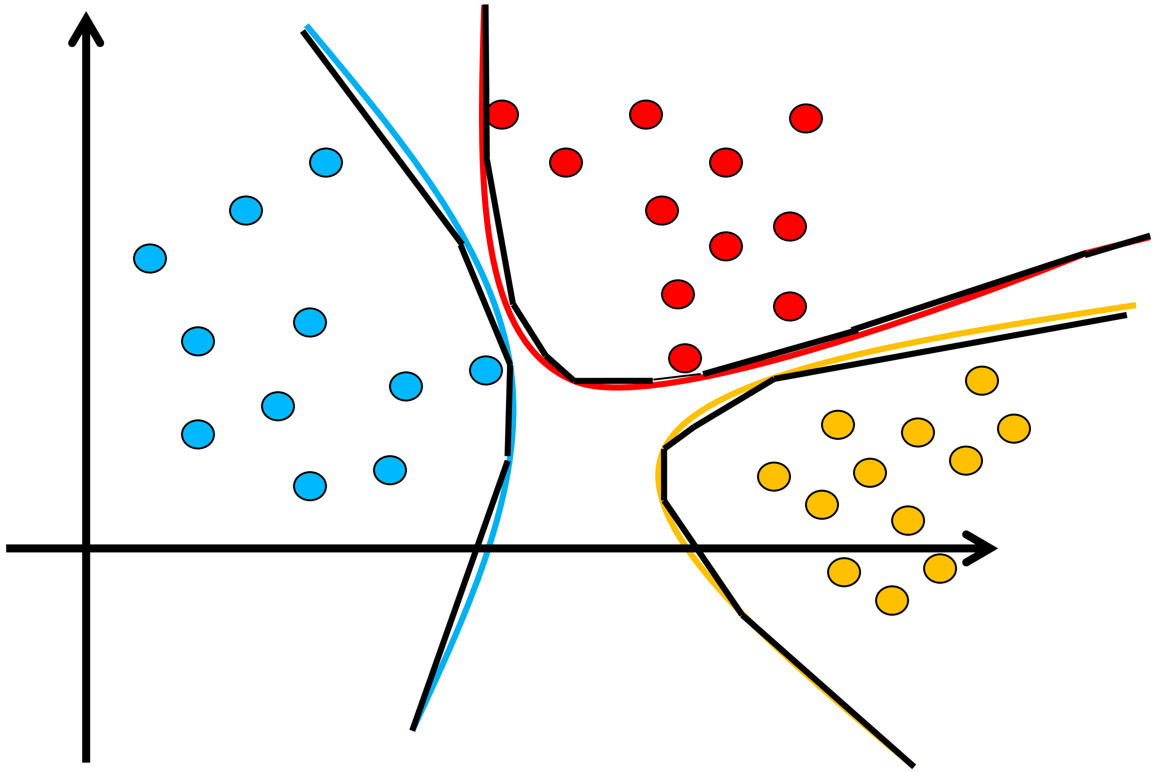


Nonlinearity

- Nonlinear approaches:
  - Piecewise linear methods: approximate the nonlinear functions with multiple linear ones

Nonlinearity

- Nonlinear approaches:
    - Piecewise linear methods

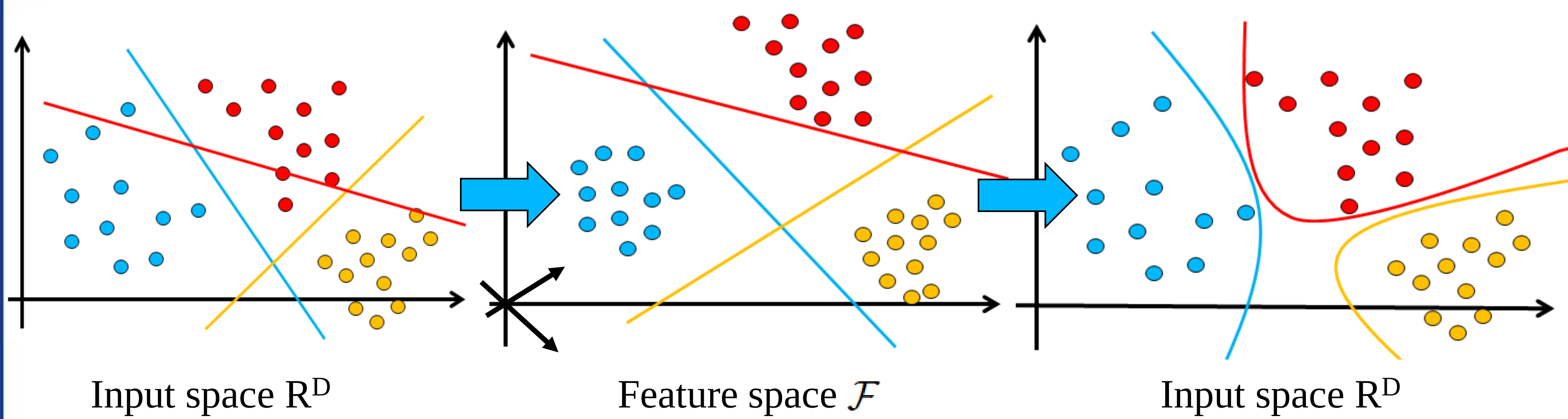


Nonlinearity

- Nonlinear approaches:
  - Piecewise linear methods
  - Kernel-based methods: Map the data (usually nonlinearly) to a feature space where the problem is easier to be solved

Nonlinearity

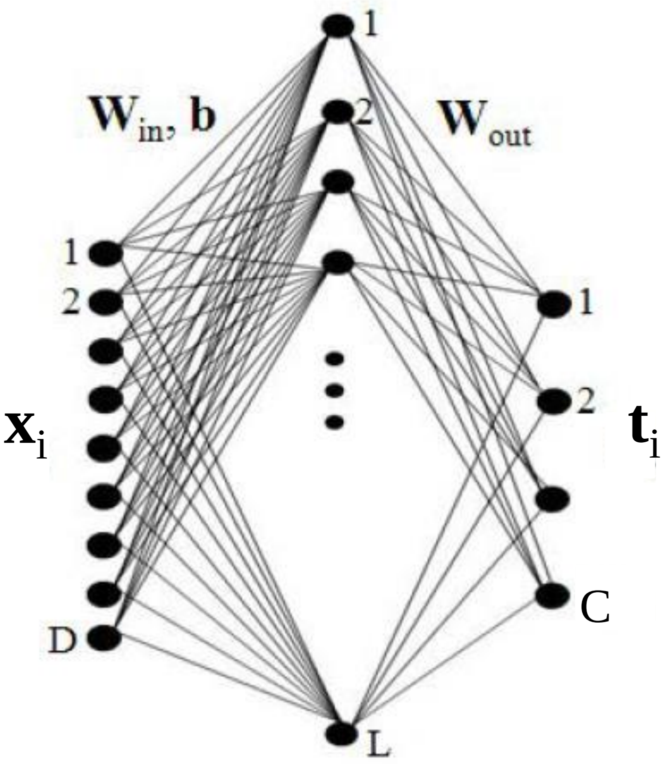
- Nonlinear approaches:
  - Piecewise linear methods
  - Kernel-based methods



Nonlinearity

- Nonlinear approaches:
  - Piecewise linear methods
  - Kernel-based methods
  - Neural Networks

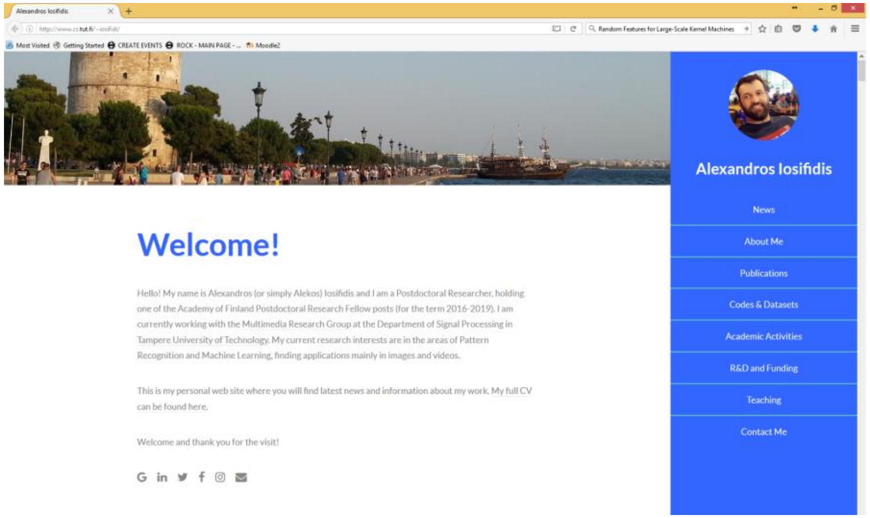
- NN with one hidden layer
- Samples $\mathbf{x}_i$ are (nonlinearly) mapped to a feature space $\mathbb{R}^L$
- The use of multiple hidden layer is also possible (each transforming the data representations in the previous layer to a new feature space).



Coming next!

- More on PR and ML stuff during this training:
  - Thursday 25.8
    14:00-16:00: Evolutionary Learning (Prof. Serkan Kiranyaz)
  - Friday 26.8
    9:00-12:00: Probabilistic and Statistical Learning (Myself)
    13:00-16:00: Neural Networks (Prof. Anastasios Tefas)

- If you want to find out more:
  http://www.cs.tut.fi/~iosifidi/



Questions ?