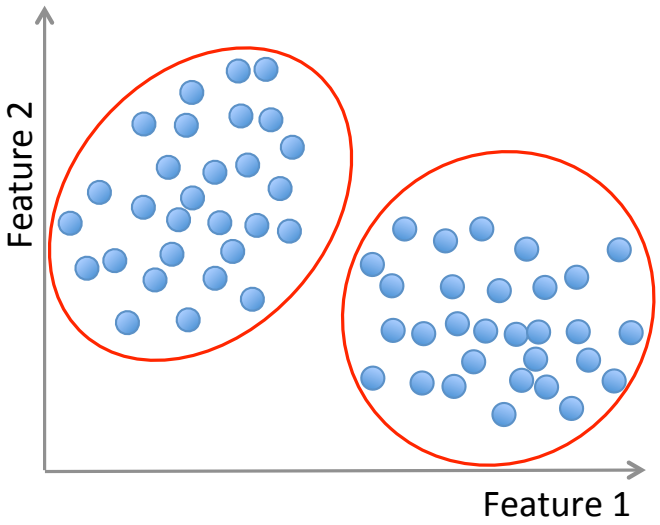
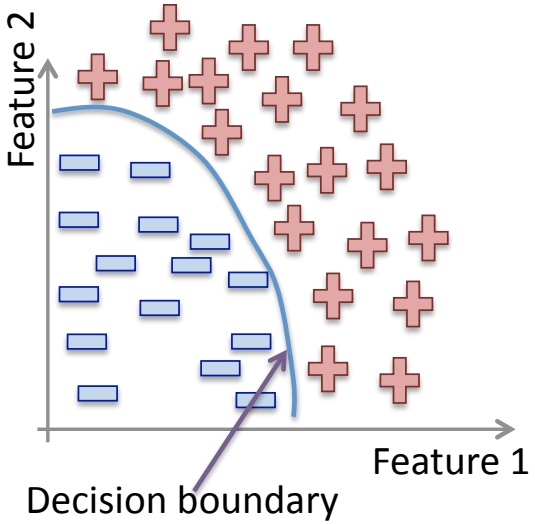
Machine Learning

Basic Concepts





Terminology

Machine Learning, Data Science, Data Mining, Data Analysis, Statistical Learning, Knowledge Discovery in Databases, Pattern Discovery.

Data everywhere!

1. **Google:** processes 24 peta bytes of data per day.
2. **Facebook:** 10 million photos uploaded every hour.
3. **Youtube:** 1 hour of video uploaded every second.
4. **Twitter:** 400 million tweets per day.
5. **Astronomy:** Satellite data is in hundreds of PB.
6. ...
7. **"By 2020 the digital universe will reach 44 zettabytes..."**

   The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, April 2014.

   That's 44 trillion gigabytes!

Data types

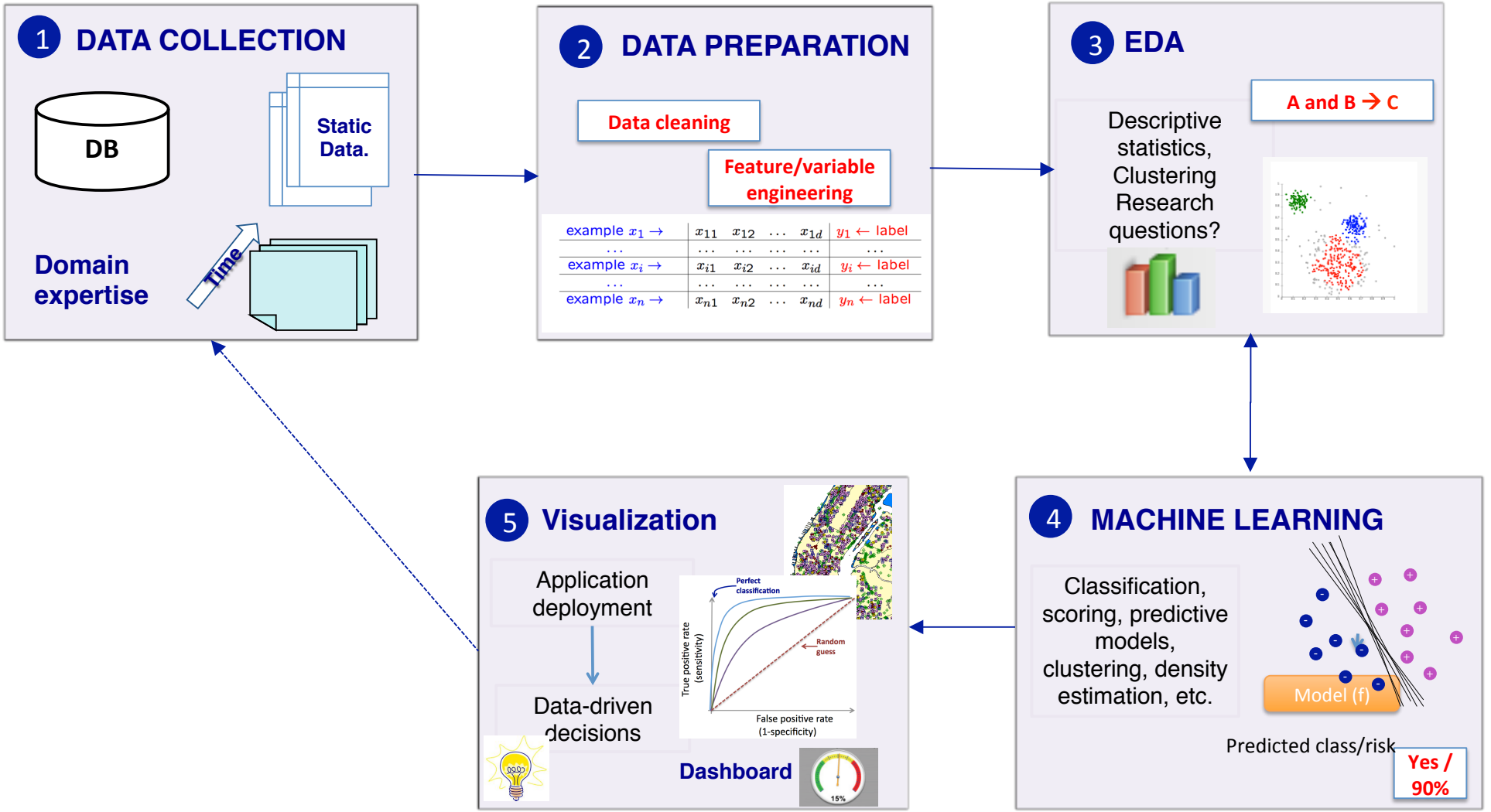
Data comes in different sizes and also flavors (types):

- ☒ **Texts**
- ☒ **Numbers**
- ☒ **Clickstreams**
- ☒ **Graphs**
- ☒ **Tables**
- ☒ **Images**
- ☒ **Transactions**
- ☒ **Videos**
- ☒ **Some or all of the above!**

Smile, we are 'DATAFIED'!

- Wherever we go, we are "datafied".
- Smartphones are tracking our locations.
- We leave a data trail in our web browsing.
- Interaction in social networks.
- Privacy is an important issue in Data Science.

The Data Science process



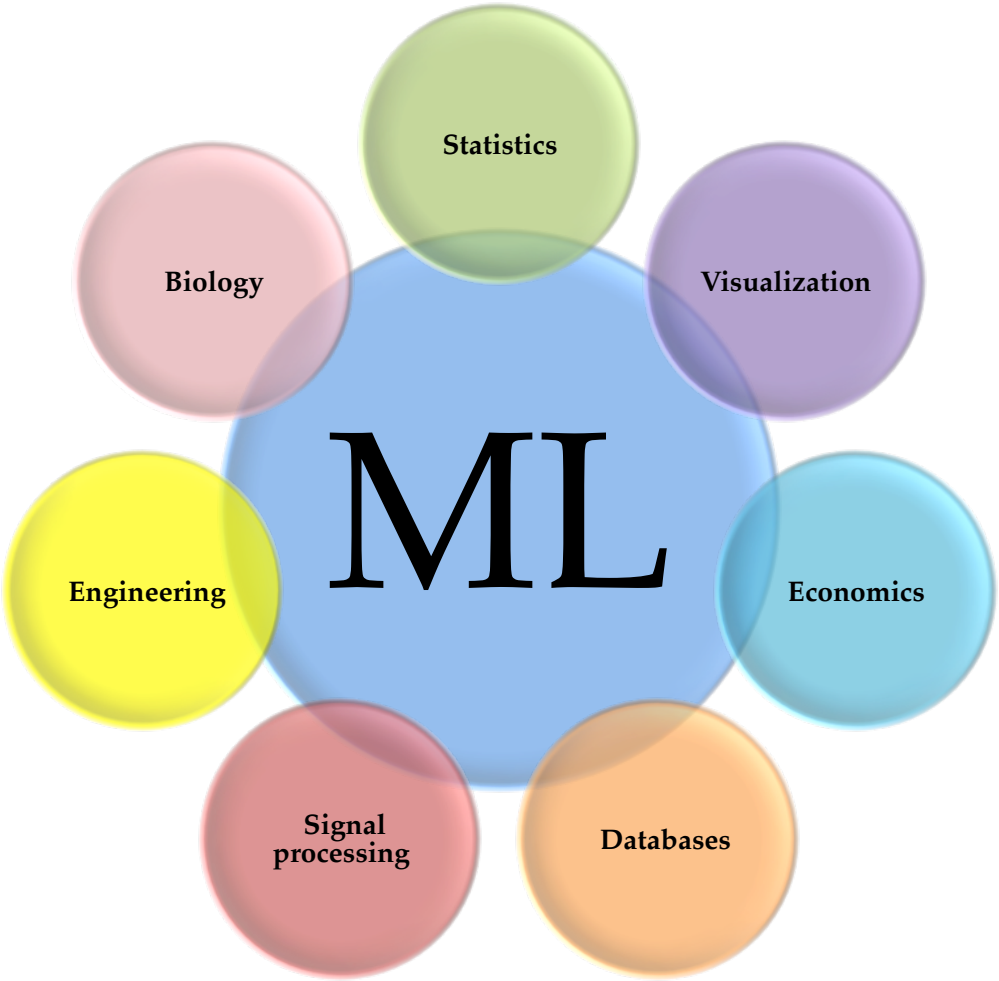
Applications of ML

- We all use it on a daily basis. Examples:

Machine Learning

- Spam filtering
- Credit card fraud detection
- Digit recognition on checks, zip codes
- Detecting faces in images
- MRI image analysis
- Recommendation system
- Search engines
- Handwriting recognition
- Scene classification
- etc...

Interdisciplinary field



ML versus Statistics

Statistics:

- Hypothesis testing
- Experimental design
- Anova
- Linear regression
- Logistic regression
- GLM
- PCA

Machine Learning:

- Decision trees
- Rule induction
- Neural Networks
- SVMs
- Clustering method
- Association rules
- Feature selection
- Visualization
- Graphical models
- Genetic algorithm

http://statweb.stanford.edu/~jhf/ftp/dm-stat.pdf

Machine Learning definition

"How do we create computer programs that improve with experience?"

Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

Machine Learning definition

"How do we create computer programs that improve with experience?"

Tom Mitchell

http://videolectures.net/mlas06_mitchell_itm/

"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. "

Tom Mitchell. Machine Learning 1997.

Supervised vs. Unsupervised

Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in \mathbb{R}^d$ and y_i is the label.

example $x_1 \rightarrow$	x_{11}	x_{12}	$\ldots$	x_{1d}	$y_1 \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_i \rightarrow$	x_{i1}	x_{i2}	$\ldots$	x_{id}	$y_i \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_n \rightarrow$	x_{n1}	x_{n2}	$\ldots$	x_{nd}	$y_n \leftarrow$ label

Supervised vs. Unsupervised

Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in \mathbb{R}^d$ and y_i is the label.

example $x_1 \rightarrow$	x_{11}	x_{12}	$\ldots$	x_{1d}	$y_1 \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_i \rightarrow$	x_{i1}	x_{i2}	$\ldots$	x_{id}	$y_i \leftarrow$ label
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
example $x_n \rightarrow$	x_{n1}	x_{n2}	$\ldots$	x_{nd}	$y_n \leftarrow$ label

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
...				
fruit n	...	...	...	...

Supervised vs. Unsupervised

fruit	length	width	weight	label
fruit 1	165	38	172	Banana
fruit 2	218	39	230	Banana
fruit 3	76	80	145	Orange
fruit 4	145	35	150	Banana
fruit 5	90	88	160	Orange
...				
fruit n	...	...	...	...

Unsupervised learning:

Learning a model from **unlabeled** data.

Supervised learning:

Learning a model from **labeled** data.

Unsupervised Learning

Training data: "examples" x.

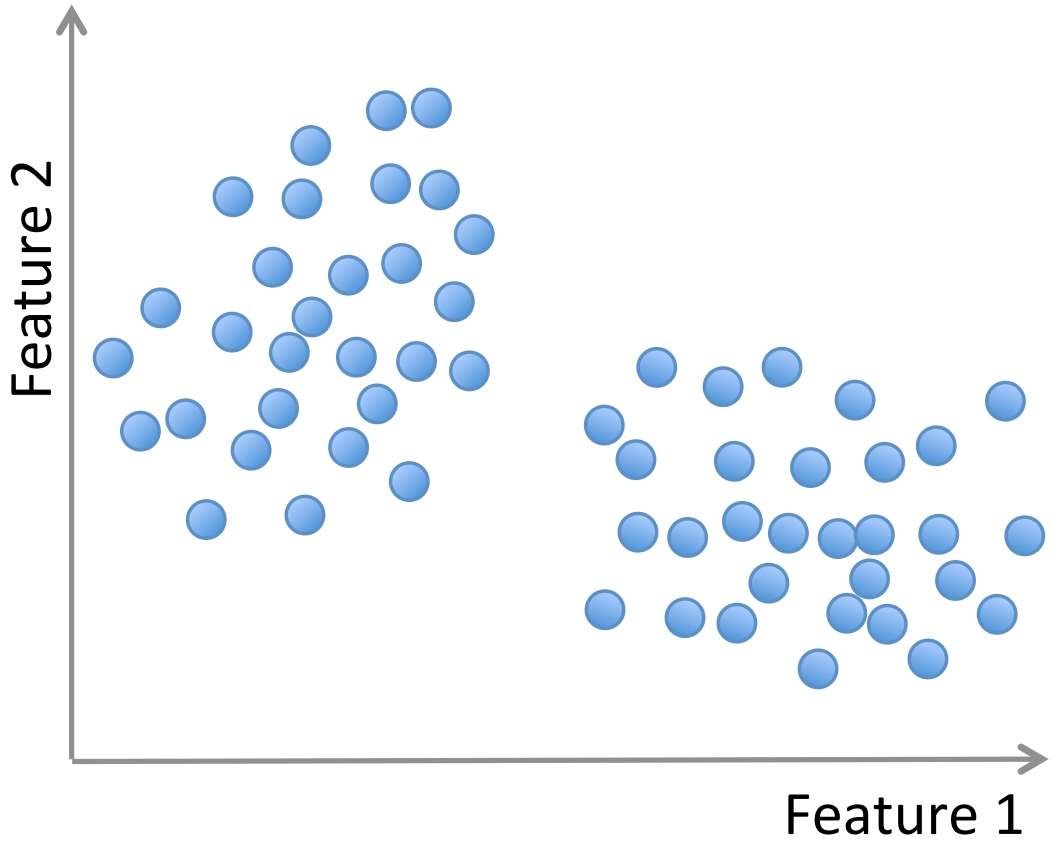
$$x_1, \ldots, x_n, \; x_i \in X \subset \mathbb{R}^n$$

- **Clustering/segmentation:**

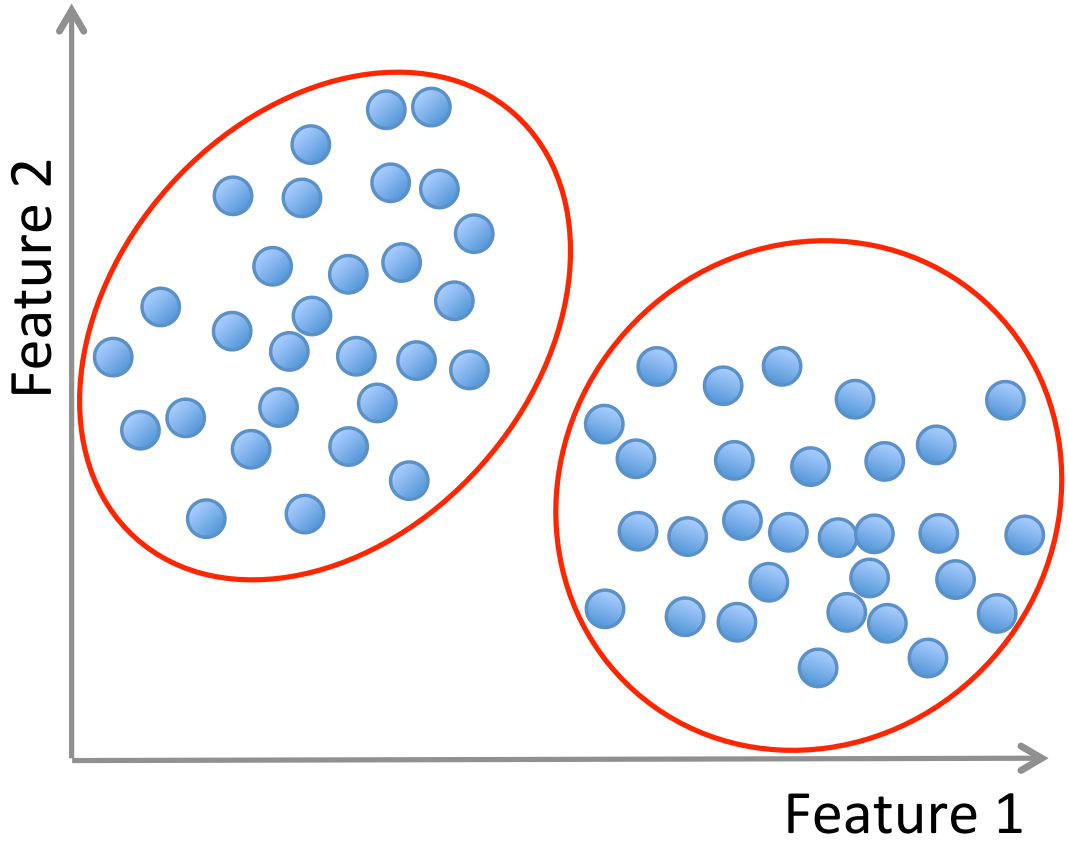
  $f : \mathbb{R}^d \longrightarrow \{C_1, \ldots C_k\}$ (set of clusters).

  Example: Find clusters in the population, fruits, species.

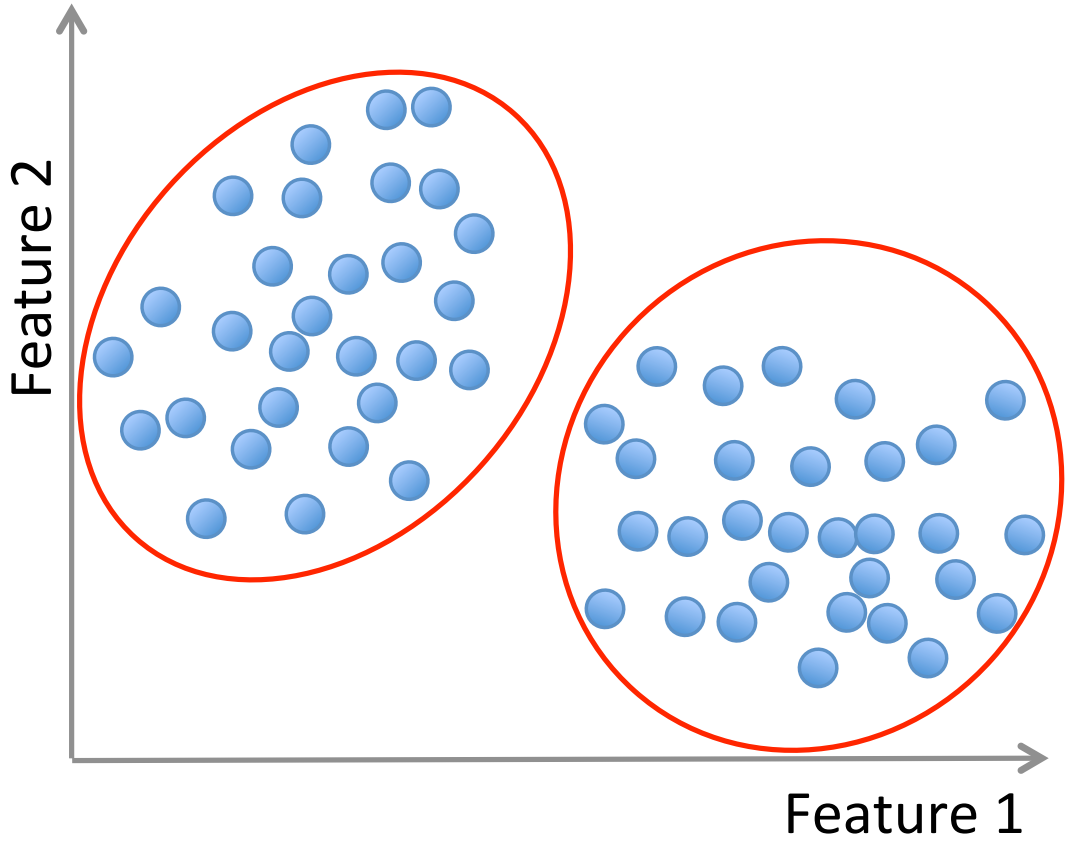
Unsupervised learning



Unsupervised learning



Unsupervised learning



Methods: K-means, gaussian mixtures, hierarchical clustering, spectral clustering, etc.

Supervised learning

Training data: "examples" x with "labels" y.

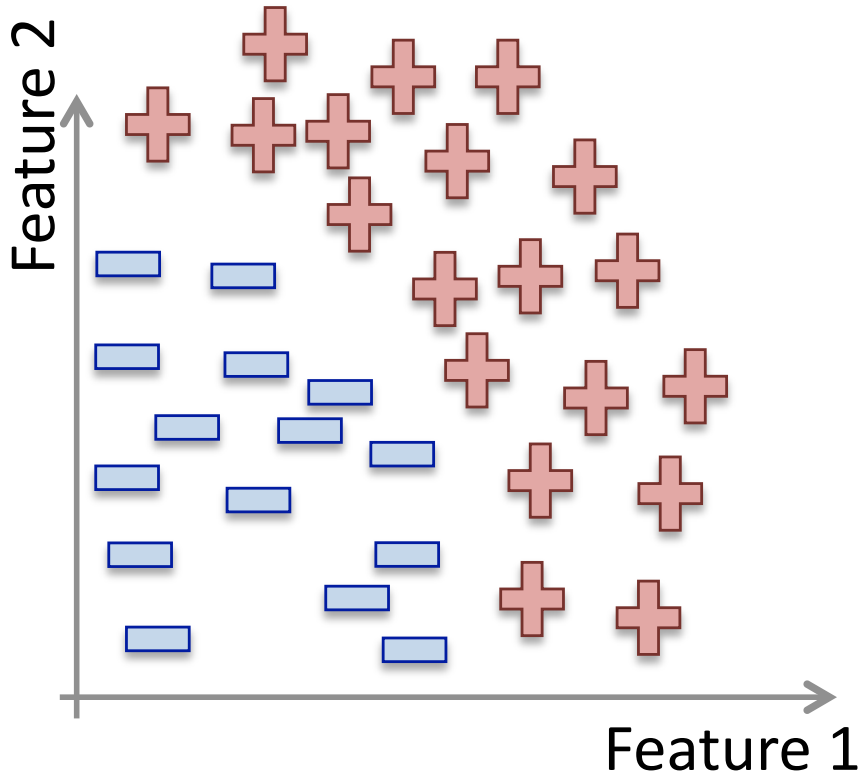
$$(x_1, y_1), \ldots, (x_n, y_n) \ / \ x_i \in \mathbb{R}^d$$

- **Classification:** y is discrete. To simplify, $y \in \{-1, +1\}$

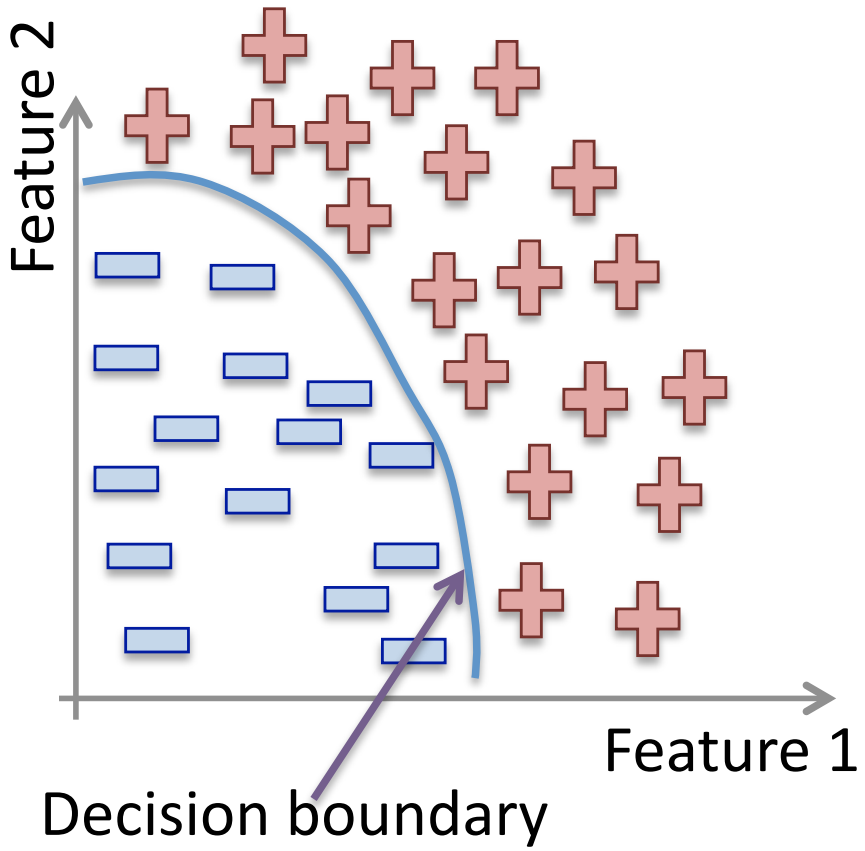
$f : \mathbb{R}^d \longrightarrow \{-1, +1\}$ f is called a **binary classifier**.

Example: Approve credit yes/no, spam/ham, banana/orange.

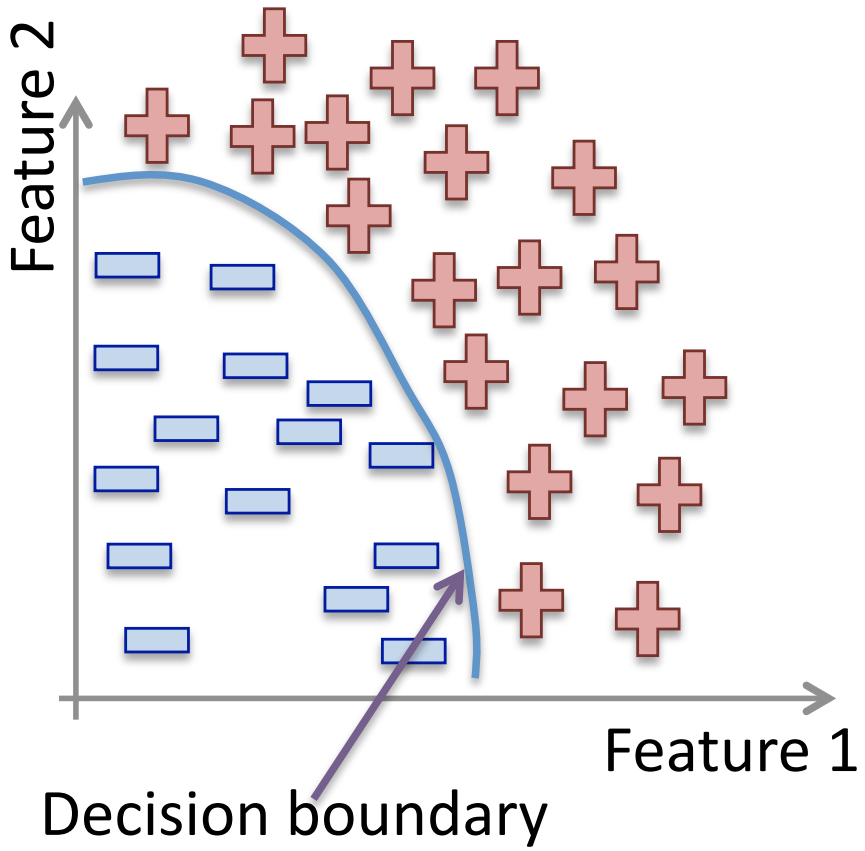
Supervised learning



Supervised learning



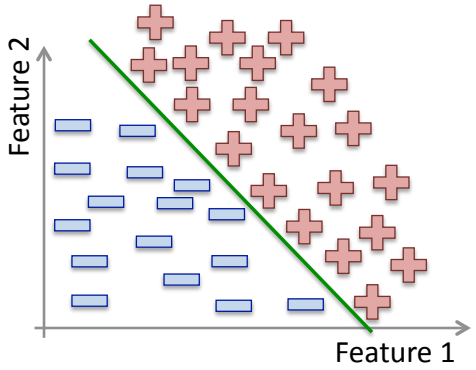
Supervised learning

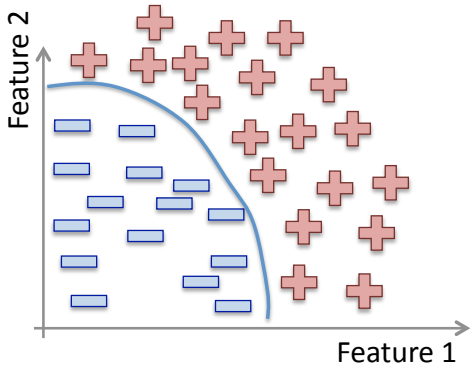


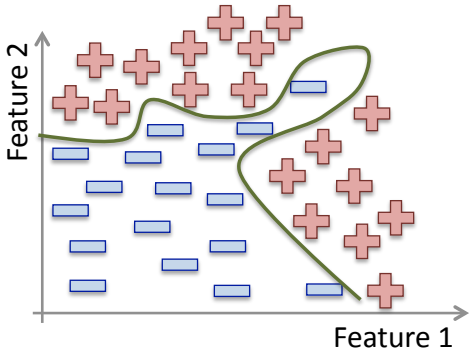
Methods: Support Vector Machines, neural networks, decision trees, K-nearest neighbors, naive Bayes, etc.

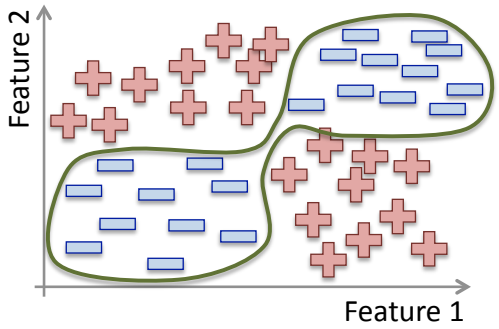
Supervised learning

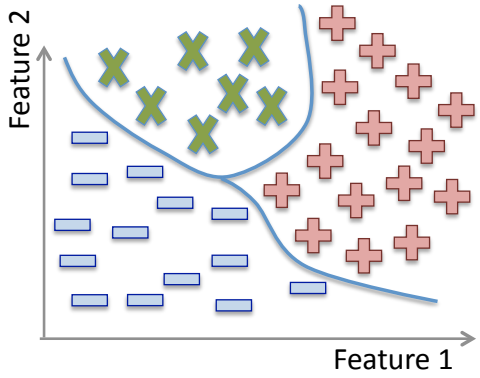
Classification:





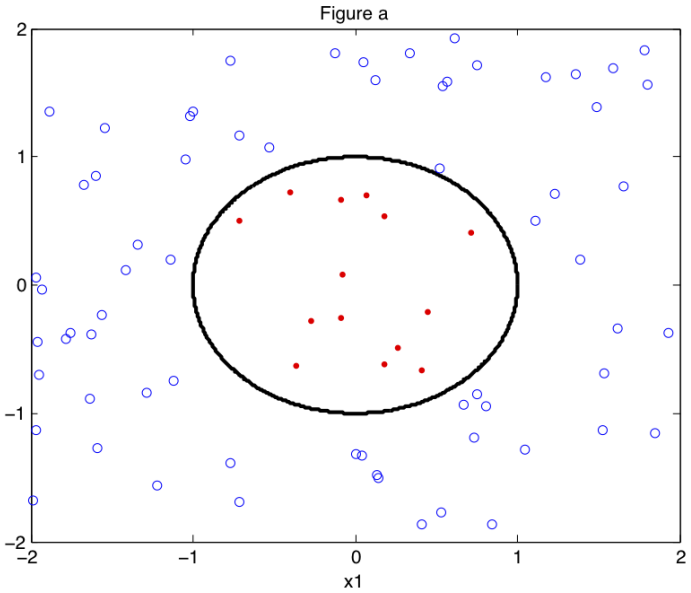


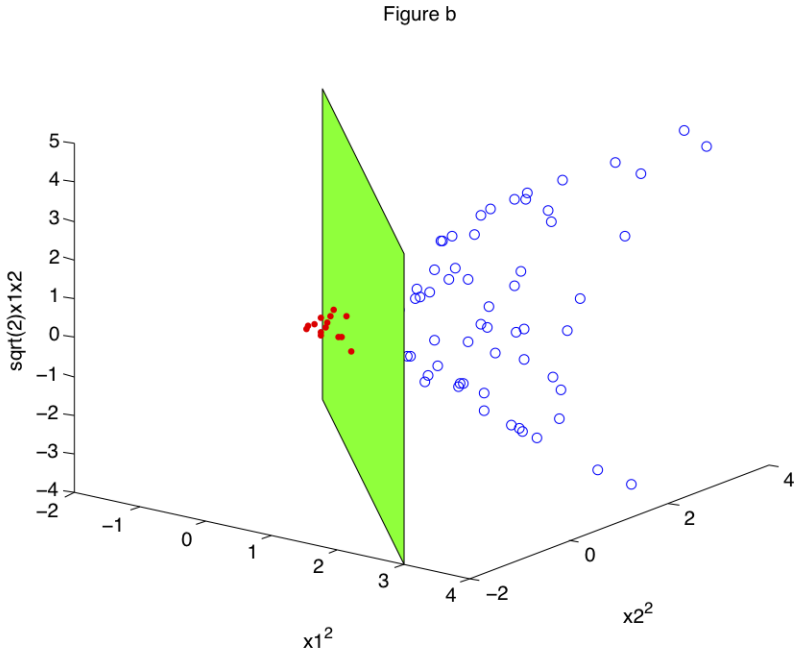




Supervised learning

Non linear classification





Supervised learning

Training data: "examples" x with "labels" y.

$$(x_1, y_1), \ldots, (x_n, y_n) \ / \ x_i \in \mathbb{R}^d$$

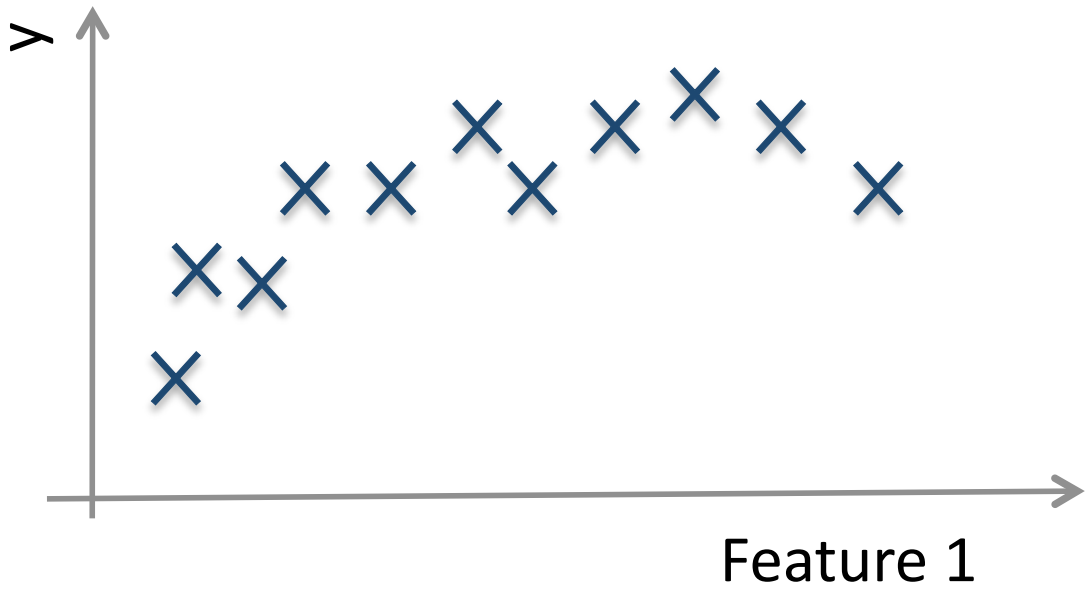
- **Regression:** y is a real value, $y \in \mathbb{R}$

  $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ f is called a **regressor**.

  Example: amount of credit, weight of fruit.

Supervised learning

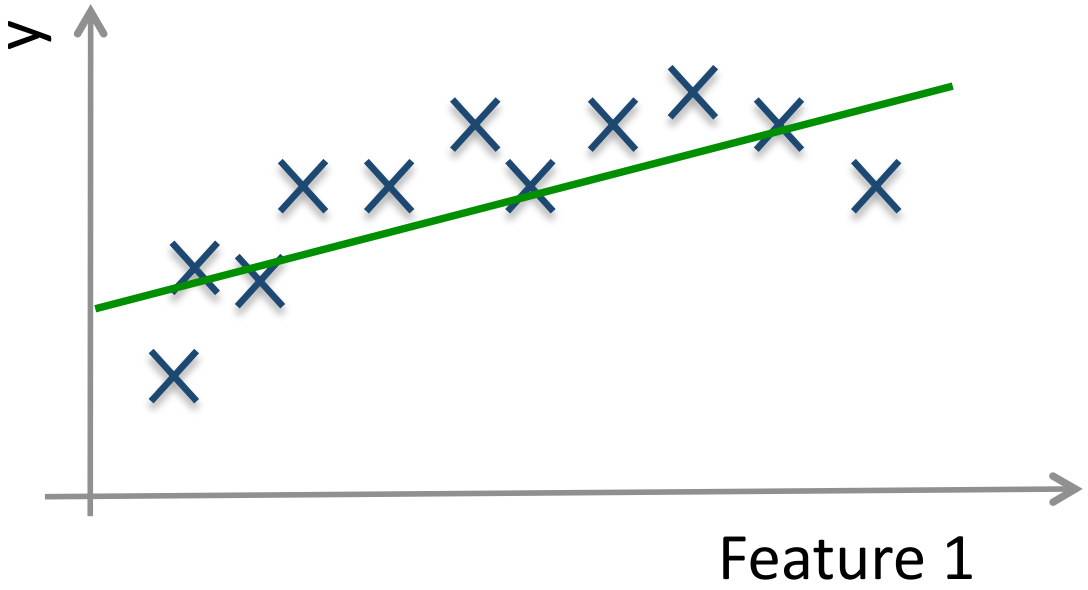
Regression:



Example: Income in function of age, weight of the fruit in function of its length.

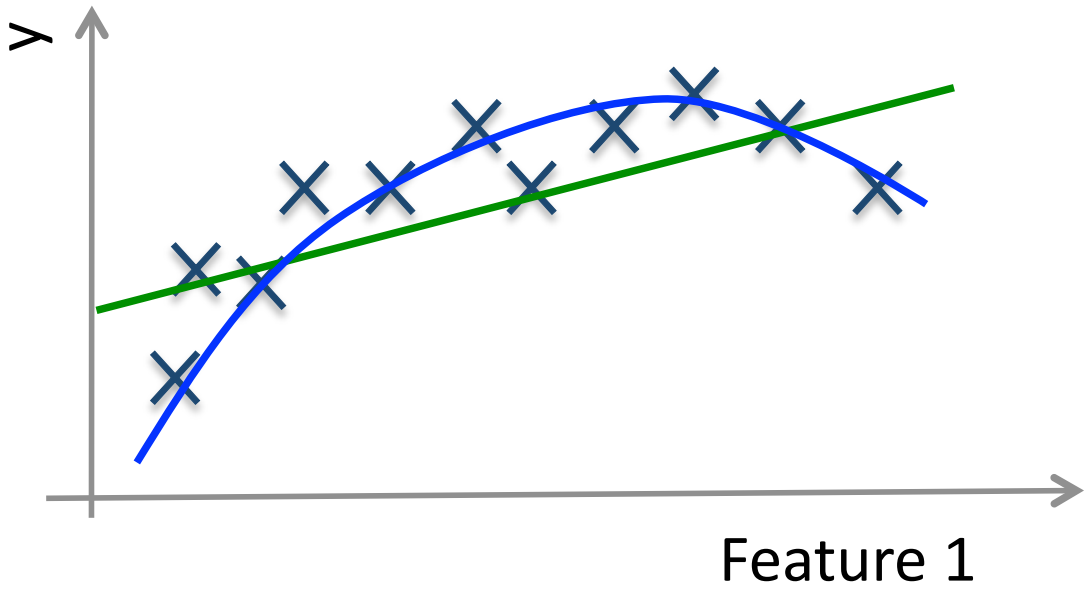
Supervised learning

Regression:



Supervised learning

Regression:



Supervised learning

Regression: