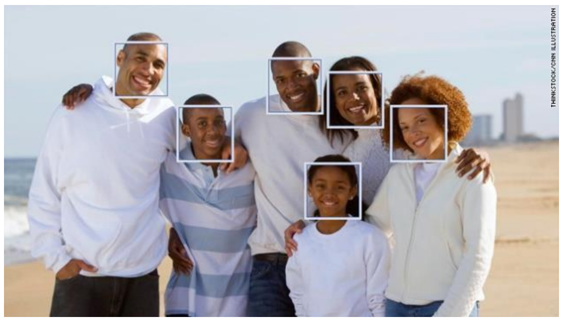
Introduction to Machine Learning

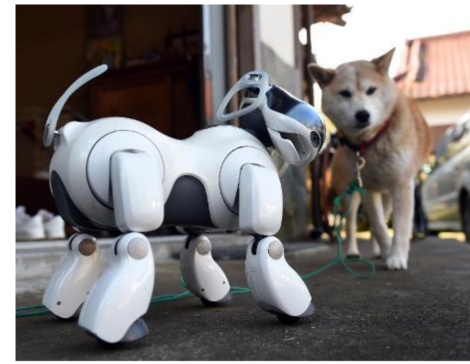
Aidos Sarsembayev, IITU, 2018

Applications of ML (and its branches)



source: cnn.com

source: kotaku.com



source: phys.org

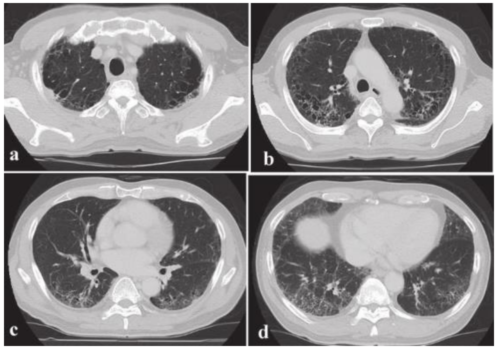
source: techinasia.com

source: 123rf.com

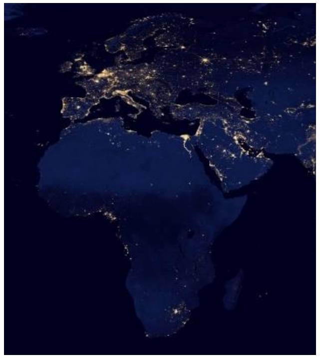
source: kvrwebtech.com

Applications of ML (and its branches)

source: techfunnel.com



source: researchgate.net



source: geology.com

source: britannica.com

Why ML is the future?

adaptivecomputing.com

Why ML is the future?

- Since the dawn of time…
- Up till 2005…
- Humans had created…
- 130 exabytes of data

- Now let’s imagine how big is exabyte?

Why ML is the future?

- ‘A’ a single letter is one byte only
- A thousand letters (‘A’ x 1000) is a single page of a small book, equals to 1 Kb
- If we multiply it by thousand again, we will get a 500 double-sided pages book. It’s only 1 Mb.

Why ML is the future?

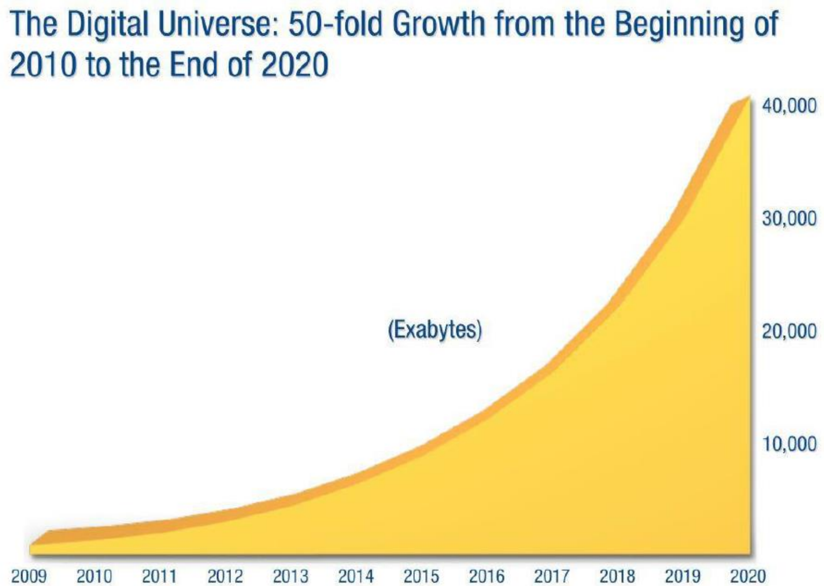
- Let’s multiply again, and we’ll get 1 Gb. Human’s genome is 720 Mb. You can easily fit it in 1 Gb.
- Let’s multiply again, and we’ll get 1 Tb. You can film ones whole life on HD camera and fit it in.
- Amazon rain forest has about 1.4 billion acres of trees. Every acre has ~500 trees. This is around 700 billion trees. If you hypothetically chop them down, put them into paper and fill every single their paper with letters, it will give you from 1 to 2 terabytes of data.

Why ML is the future?

- Now, if you multiply by 1000 again, this will give you 1 Exabyte.
- Again, by 2005 the humanity produced ~130 Exabytes of data.
- By 2010 it reached 1200 Exabytes…
- By 2015 it became **!7900!** Exabytes!

Why ML is the future?

- By 2020 it will be 40900 Exabytes!!!



This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

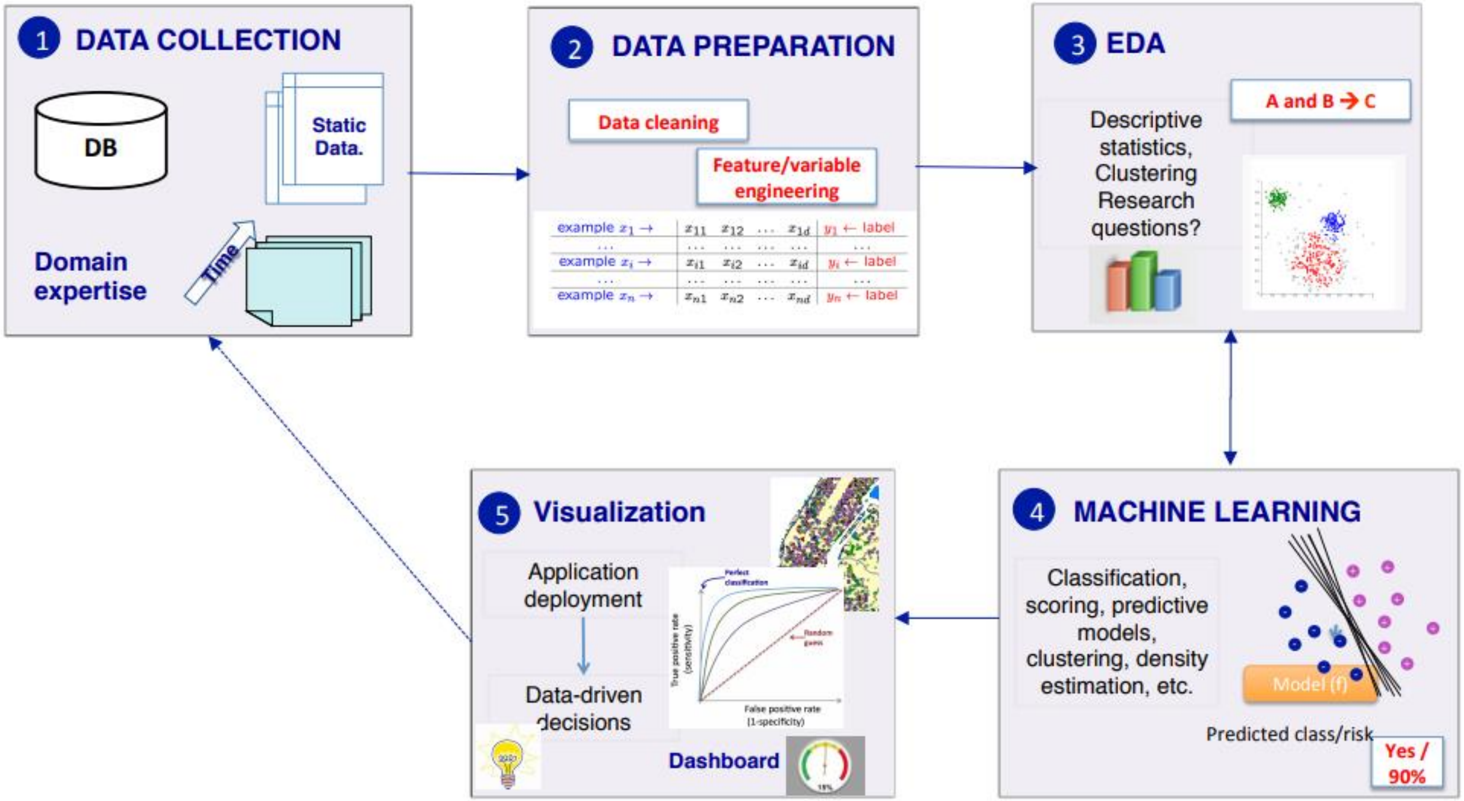
What kinds of data is it?

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

How do we tackle these data?

- Text – NLP
- Numbers – Regression, Deep Learning algorithms,
- Images, videos - Deep Learning (ConvNets) algorithms
- Clickstreams, transactions – artificial neural networks
- etc.

How does the ML process look like?



- “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. ”

Supervised vs. Unsupervised

- Given: Training data: $(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in R^d$ 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: $(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in R^d$ 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

- Unsupervised learning: Learning a model from unlabeled data.
- Supervised learning: Learning a model from labeled data.

Unsupervised Learning

- Training data:“examples” x.

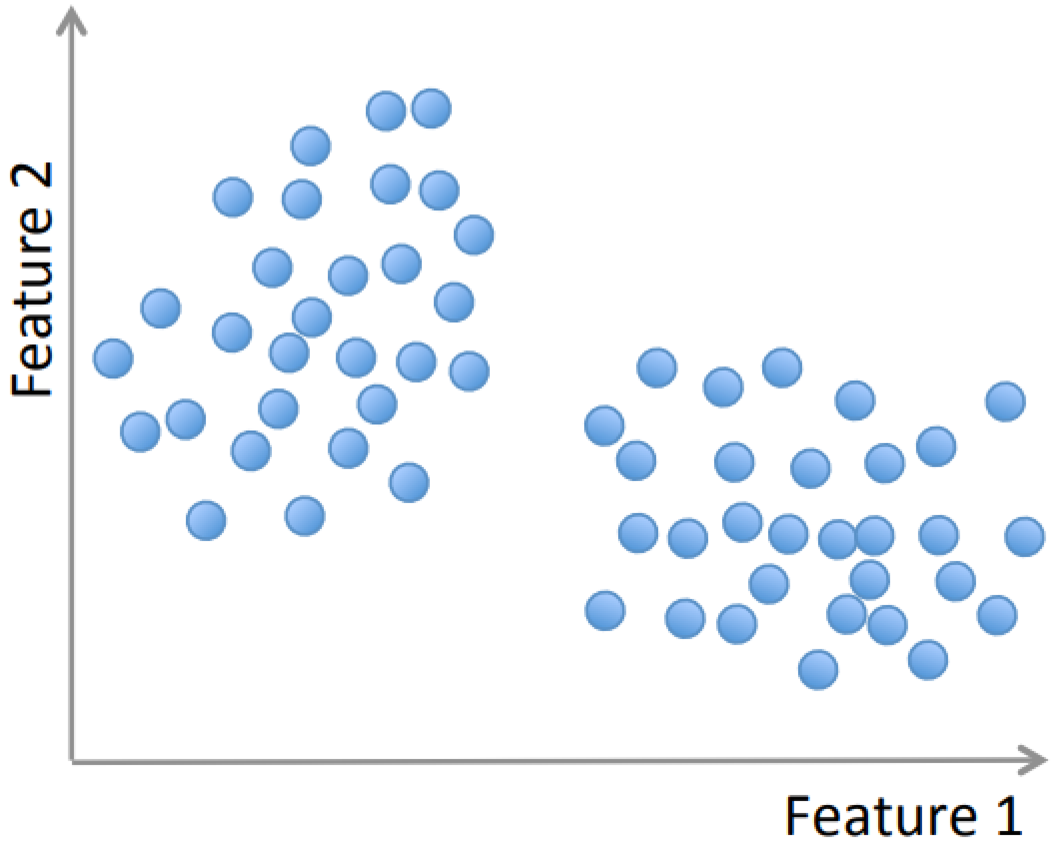
$x_1, \ldots, x_n, x_i \in X \subset R^n$

Clustering/segmentation:

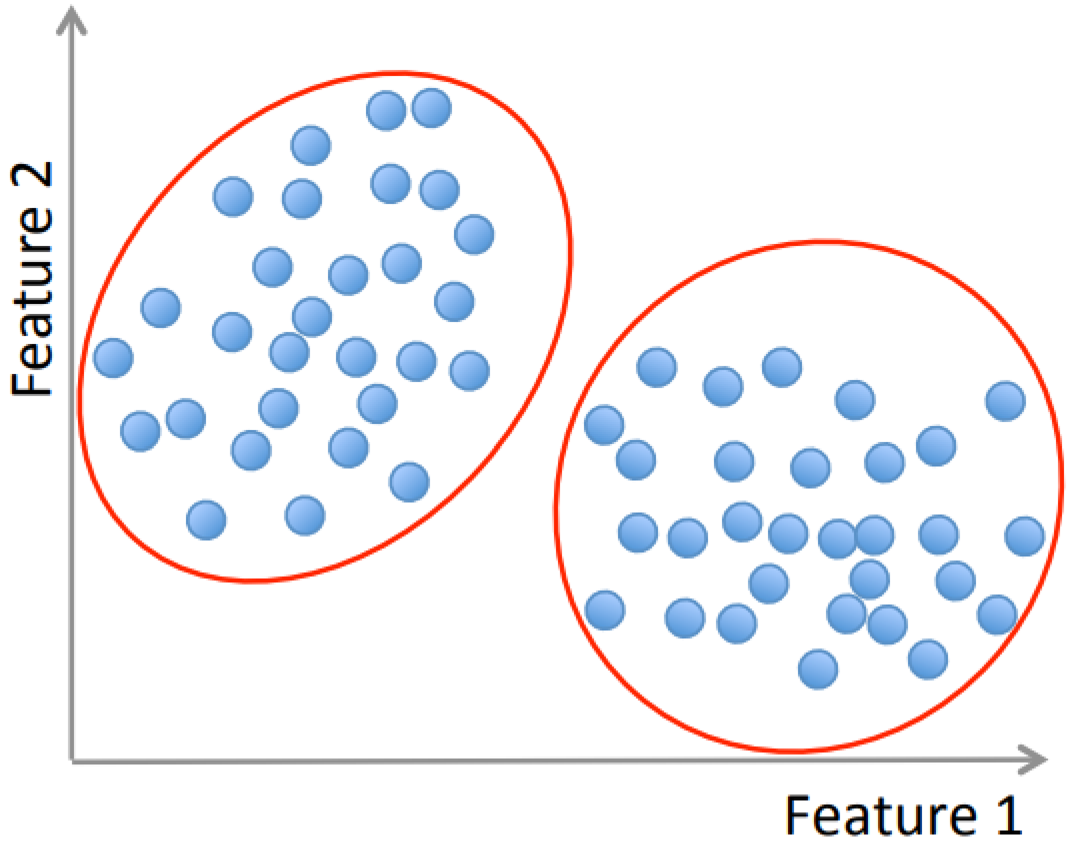
$f : R^d \longrightarrow \{C_1, \ldots C_k\}$ (set of clusters).

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

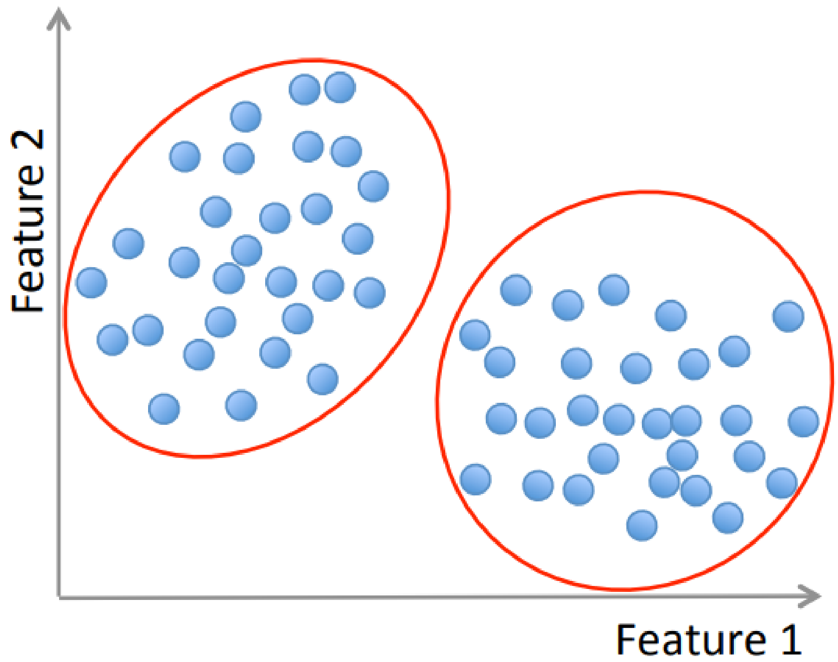
Unsupervised learning



Unsupervised learning



Unsupervised learning



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

Supervised learning

- Training data: “examples” x with “labels” y.

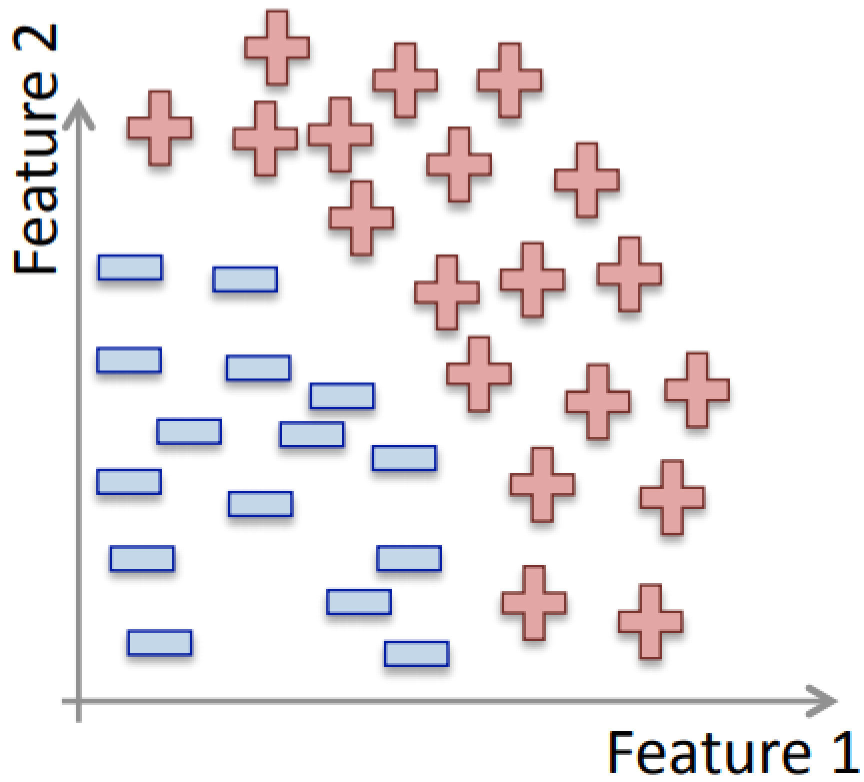
$(x_1, y_1), \ldots, (x_n, y_n)$ / $x_i \in R^d$

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

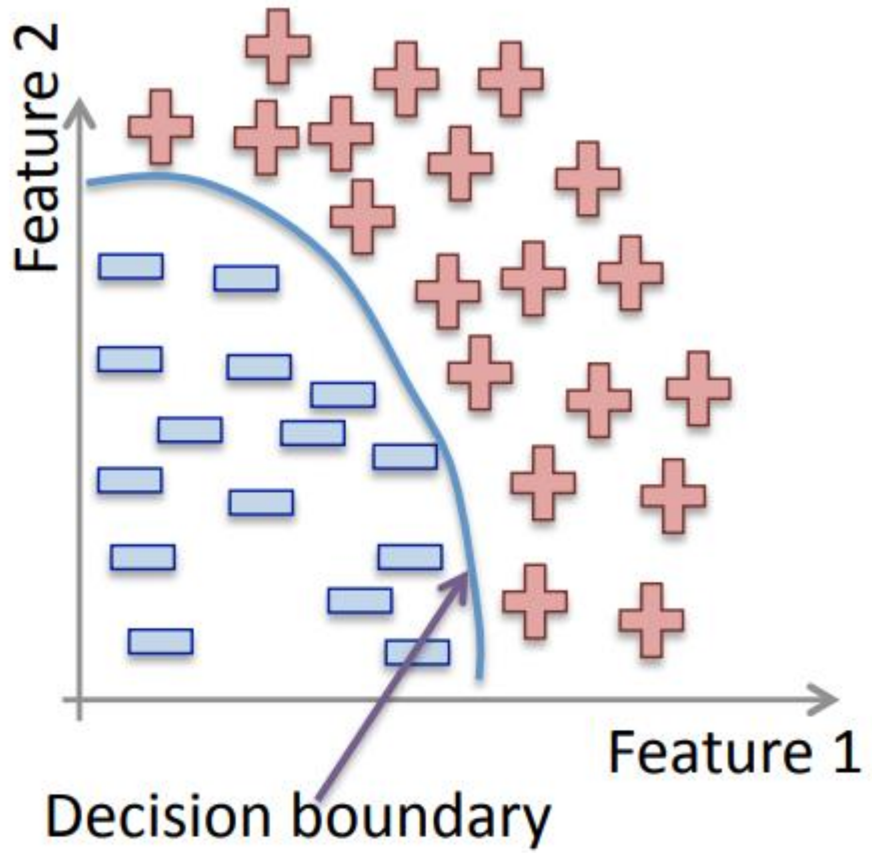
$f : R^d \longrightarrow \{-1, +1\}$ f is called a binary classifier

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

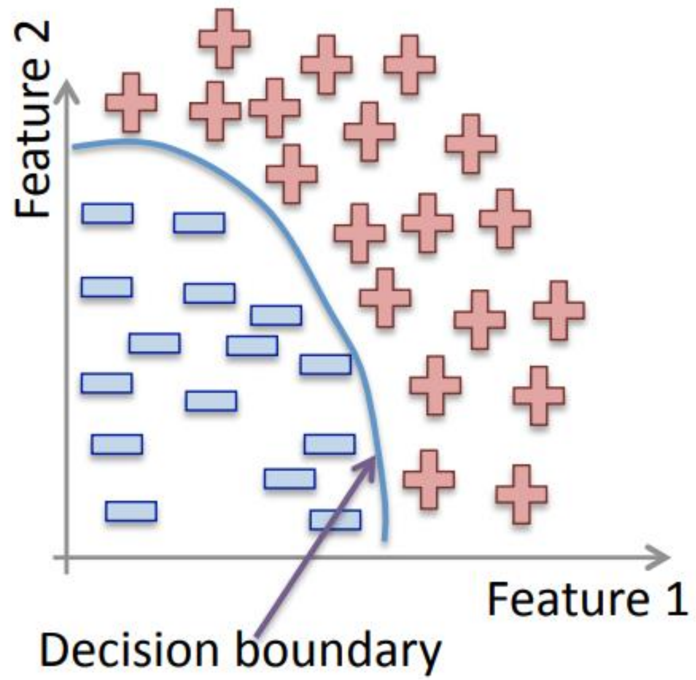
Supervised learning



Supervised learning

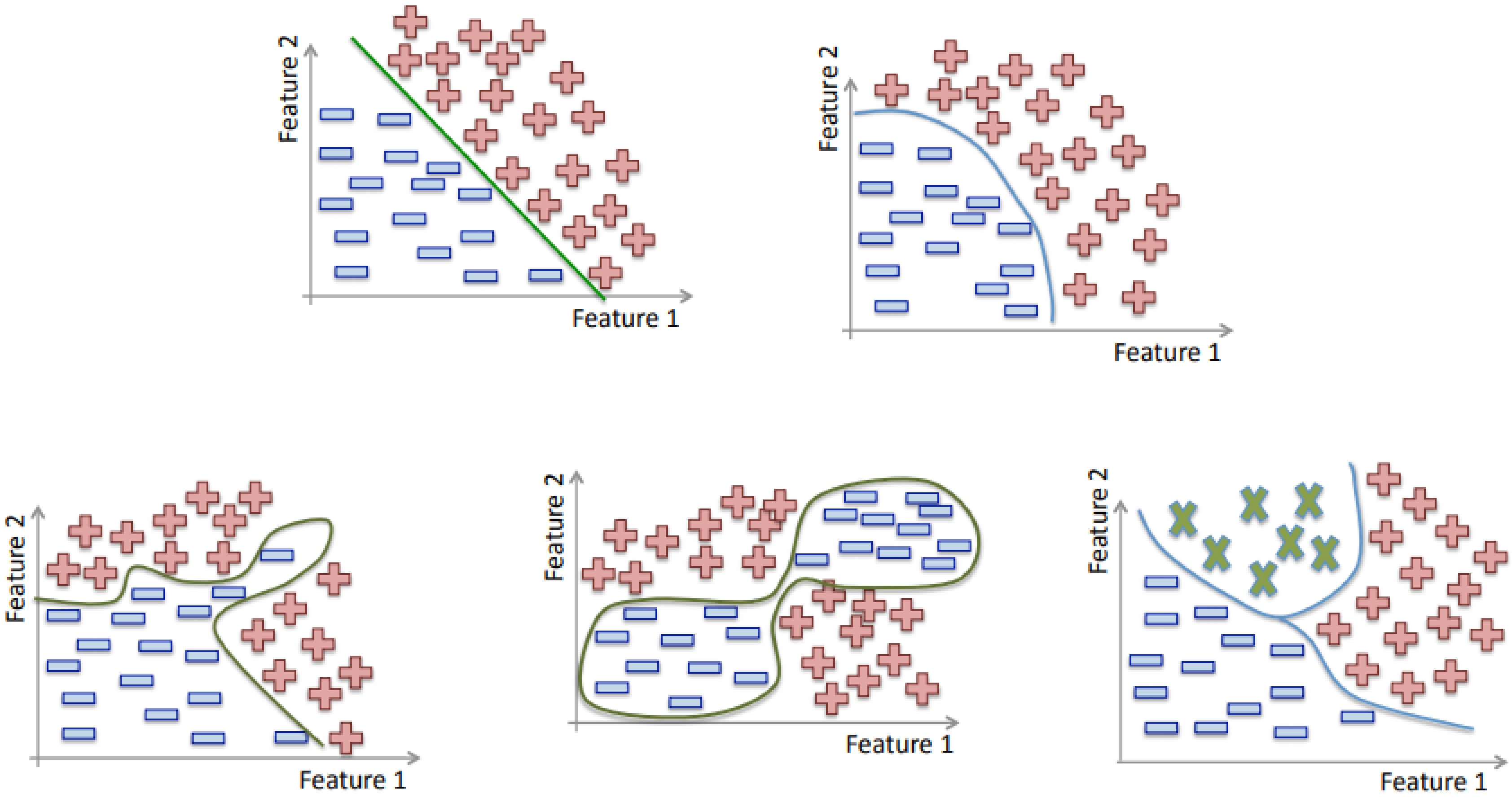


Supervised learning



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

Classification



Supervised learning

- Training data:“examples” x with “labels” y.

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

Regression: y is a real value, $y \in R$

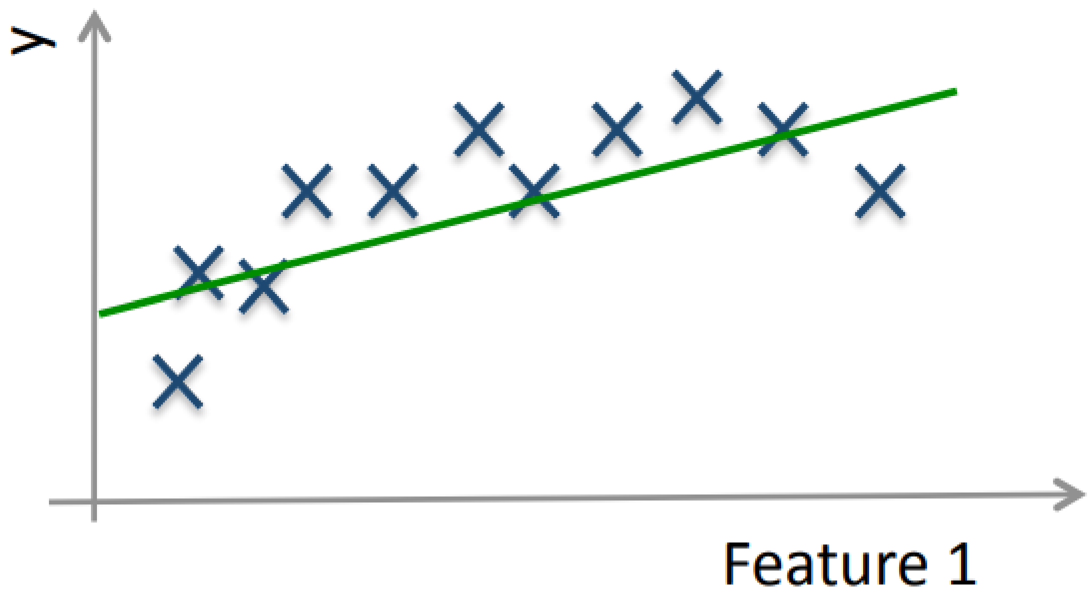
$f : R^d \rightarrow R$ f is called a regressor.

Example: amount of credit, weight of fruit.

Regression

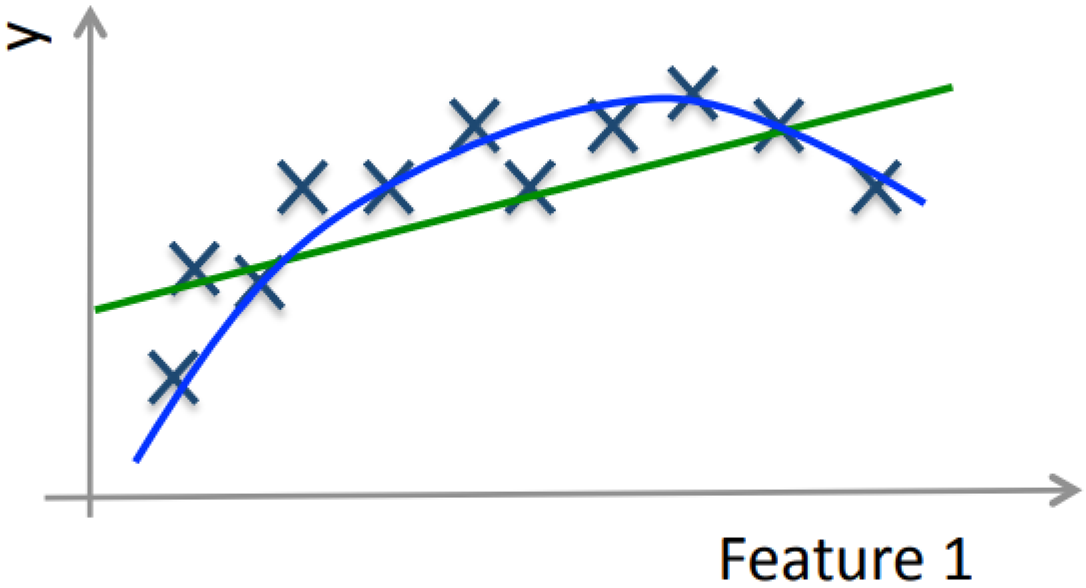
Example: Income in function of age, weight of the fruit in function of its length.

Regression



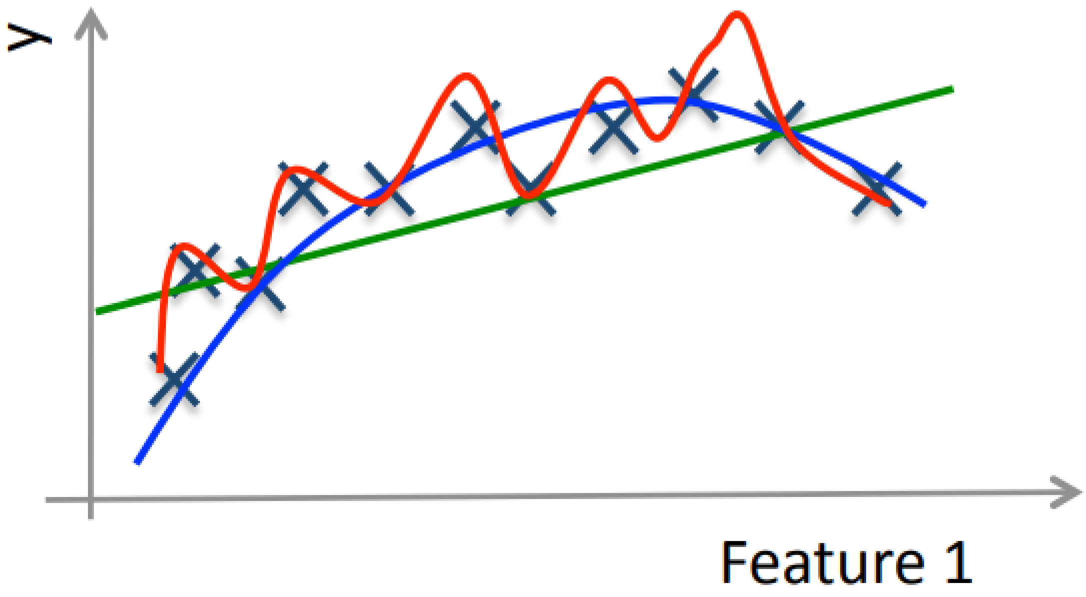
Example: Income in function of age, weight of the fruit in function of its length.

Regression



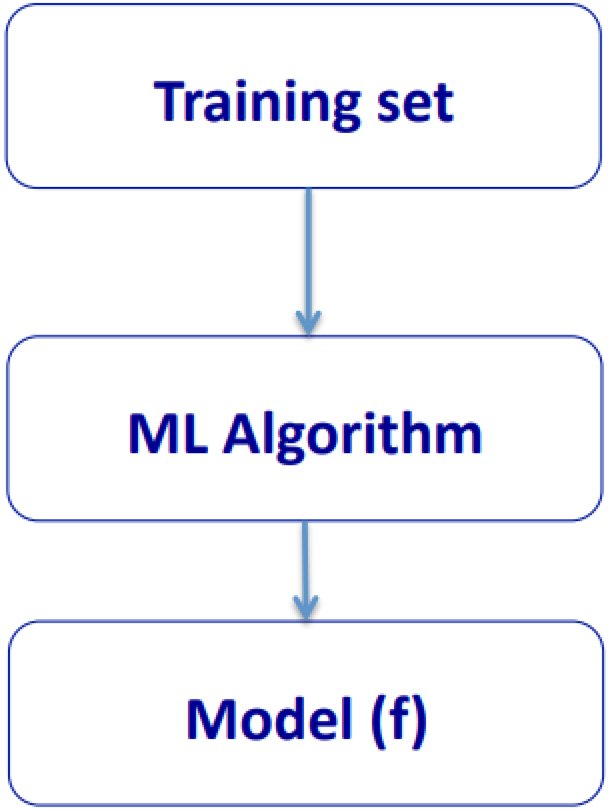
Example: Income in function of age, weight of the fruit in function of its length.

Regression

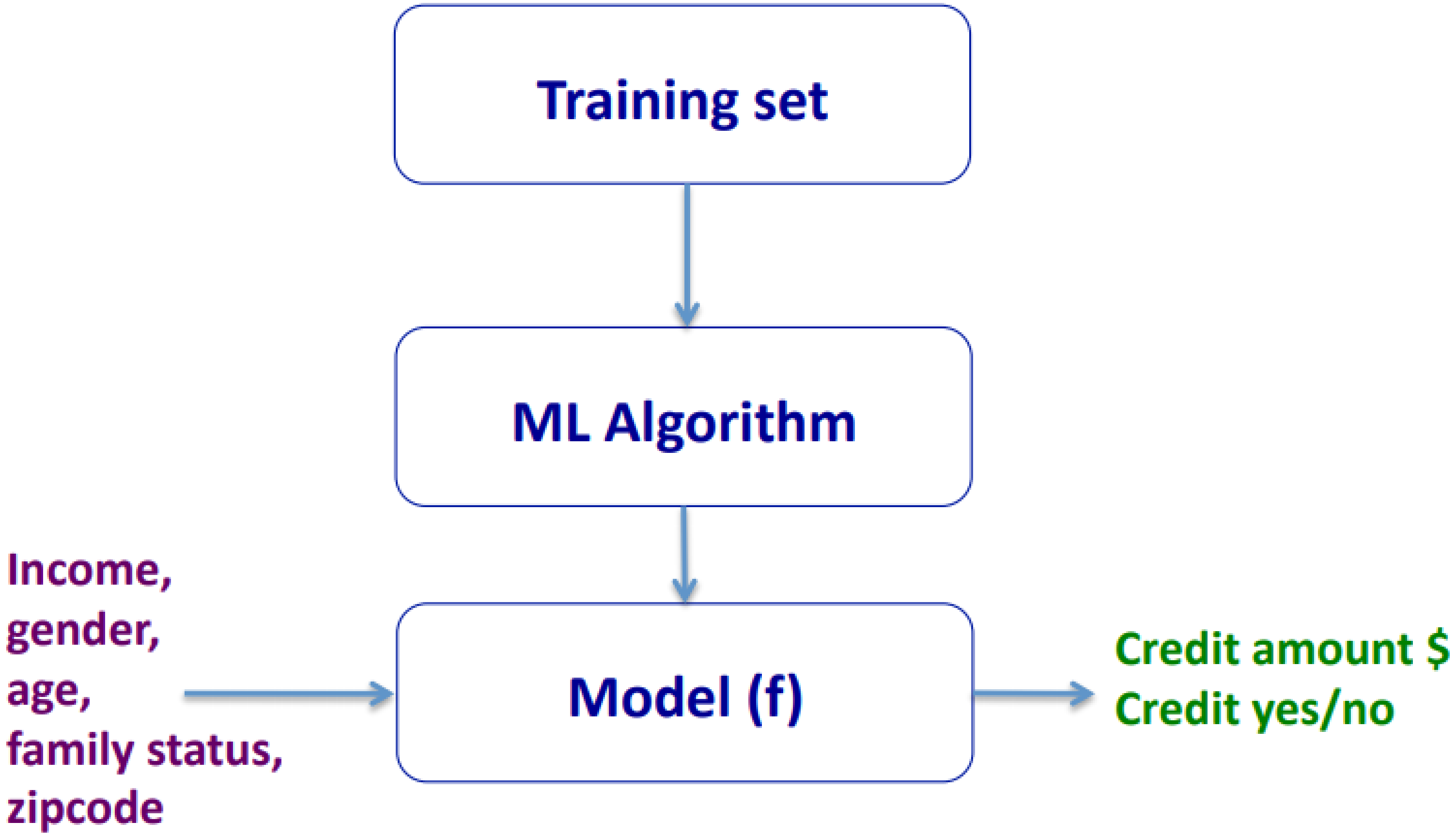


Example: Income in function of age, weight of the fruit in function of its length.

Training and Testing



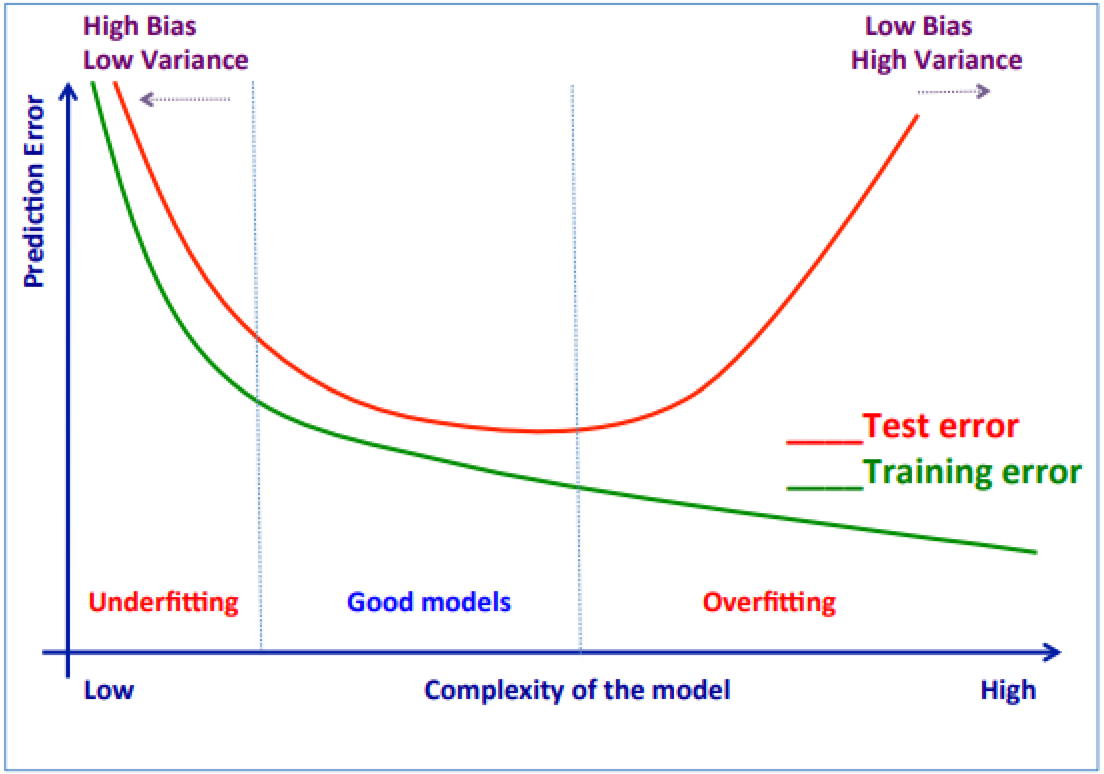
Training and Testing



Training and Testing

- We usually split our datasets into training and test sets (80/20 %%, 70/30 %% etc.)
- The effectiveness of the algorithm is calculated by accuracy and loss functions

Overfitting/underfitting



Confusion matrix

		Actual Label	
		Positive	Negative
Predicted Label	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Evaluation metrics

		Actual Label	
		Positive	Negative
Predicted Label	Positive	**True Positive (TP)**	**False Positive (FP)**
	Negative	**False Negative (FN)**	**True Negative (TN)**

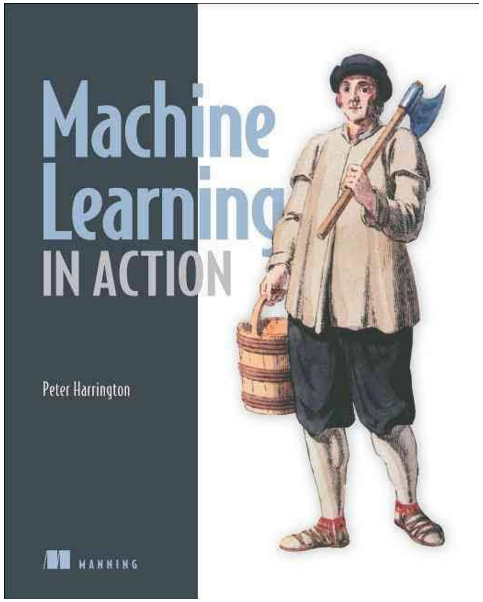
Accuracy	(TP + TN) / (TP + TN + FP + FN)	The percentage of predictions that are correct
Precision	TP / (TP + FP)	The percentage of positive predictions that are correct
Sensitivity (Recall)	TP / (TP + FN)	The percentage of positive cases that were predicted as positive
Specificity	TN / (TN + FP)	The percentage of negative cases that were predicted as negative

Literature

- Machine Learning in Action, Peter Harrington
- A Course in Machine Learning, Hal Daumé III, 2013
- Pattern Recognition and Machine Learning, Bishop, C., 2006

Дополнительная литература

- Machine Learning with R, Brett Lantz, 2015
- Python Machine Learning, Sebastian Raschka, 2015



Credits:

- most of the figures are taken from ColumbiaX online course on Machine Learning taught by Professor John W. Paisley. The course can be found on EdX platform