Introduction to Machine Learning

Lesson given by **Sébastien Piérard** in the course
"Vision 3D" (ULg, Pr. M. Van Droogenbroeck)

INTELSIG, Montefiore Institute, University of Liège, Belgium

November 30, 2011

An introductory example

- Can you decide which silhouettes are those of humans ?
- Try to write an algorithm to solve this problem !

An introductory example

- Can you decide which silhouettes are those of humans ?
- Try to write an algorithm to solve this problem !

An introductory example

Observation

Most of the tasks related to video scene interpretation are complex. A human expert can easily take the right decision, but usually without being able to explain how he does it.

Solution

Machine learning techniques are indispensable in computer science.

Outline

Outline

Machine learning techniques

Their[1] aim is to

- to build a decision rule automatically
- to be able to generalize to unseen objects
- and to speedup the decisions.

Computational cost:

- The model is learned only once.
- The model is used many times.
- Which operation should be the fastest ?

[1] In this document, we consider only "supervised" machine learning techniques.

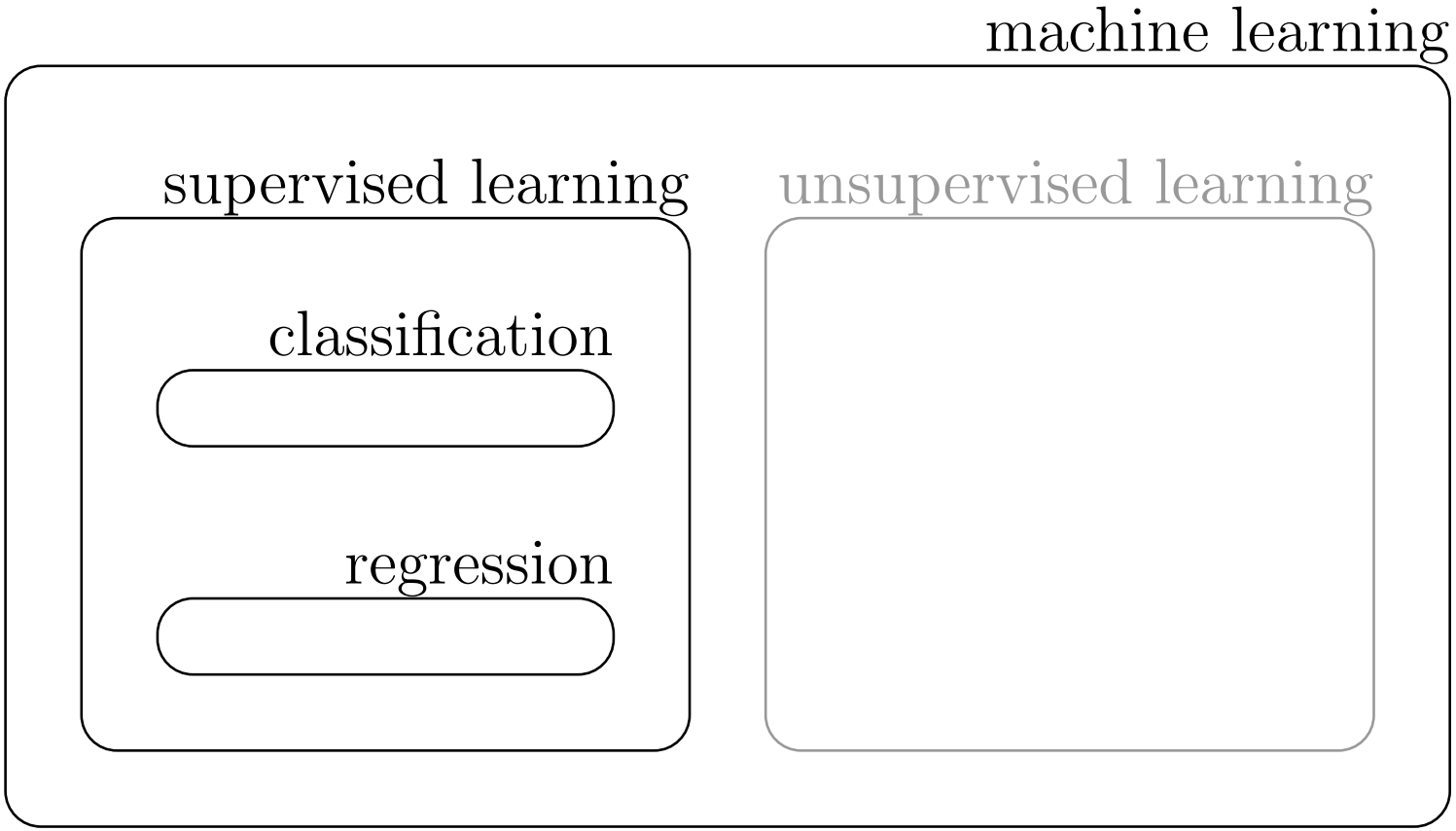
Machine learning techniques

Examples of techniques are:

- the nearest neighbors;
- the neuronal networks;
- the *ExtRaTrees* [5];
- and the *Support Vector Machines* (SVM) [2].

A good reference book on this topic is [6].

Families of machine learning methods



Applications

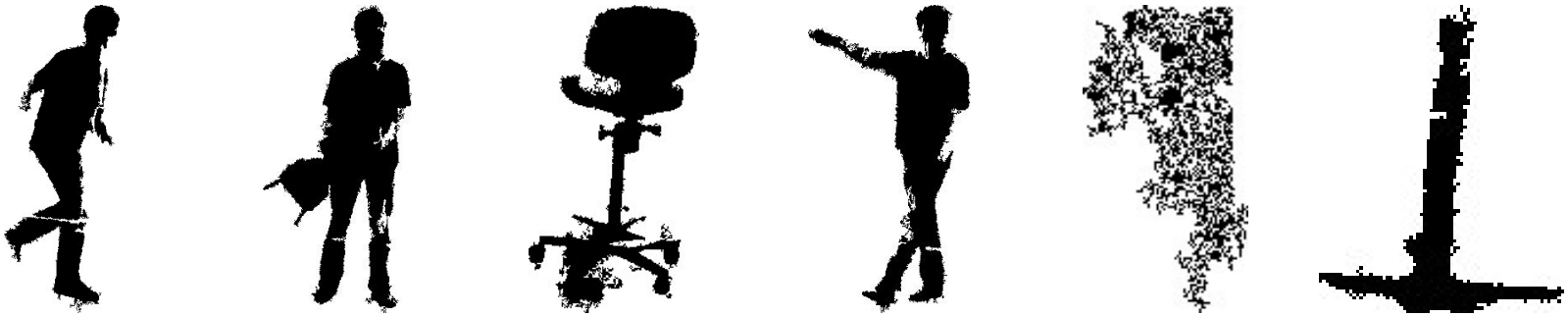
Such techniques have proven to be successful for many purposes:

- detecting people in images [3];
- recognizing them [1];
- analyzing their behavior [7];
- detector faces with software embedded on cameras [8];
- etc.

[image source: Shotton2011RealTime]

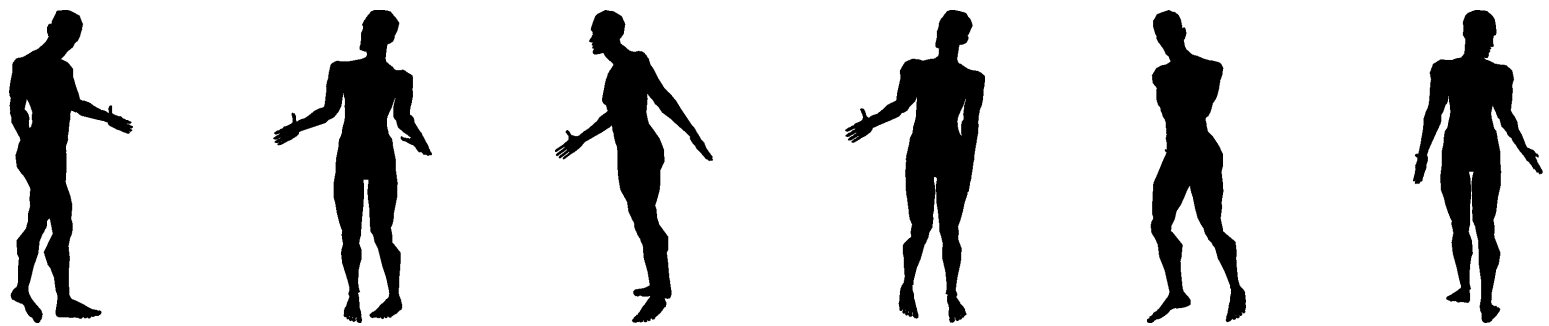
Classification *vs* regression

Example of classification



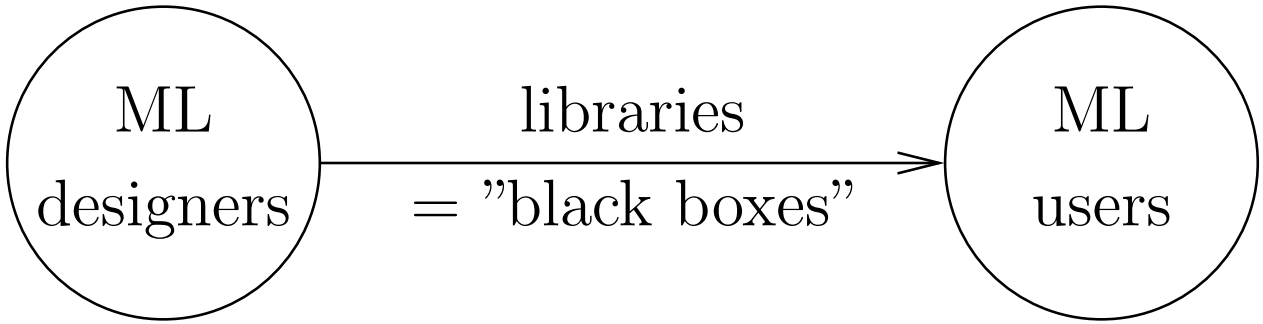
yes yes no yes no no

Example of regression



65.2° −2.0° −71.5° 15.4° −47.4° −5.5°

In practice ...



There exists a lot of machine learning libraries. For example,

- libSVM (Matlab, Java, Python, etc)
- Regression trees (C/Matlab, on Pierre Geurts's webpage)
- Neural Network Toolbox (Matlab)
- Java-ML (Java)
- Shark (C++)
- Shogun (C++)
- ...

Outline

What is learned

Example of learning database:

	x_1	x_2	x_3	x_4	x_5	y
sample 1	7.99	6.77	9.75	1.58	1.00	0
sample 2	2.24	9.51	1.14	8.00	7.66	0
sample 3	2.18	2.83	2.96	5.14	9.73	0
sample 4	8.44	7.39	4.57	4.94	2.70	1
sample 5	9.55	5.92	2.52	0.46	1.53	1
sample 6	3.32	9.13	0.50	5.07	8.22	2

$$MODEL \equiv y\left(x_1, x_2, x_3, x_4, x_5\right) = ?$$

- y is the output variable (the class)
- The samples have to be described by *attributes* x_1, x_2, ...
- The same number of attributes should be used for all samples.
- The meaning of an attribute should not depend on the sample.

Example of classification task

handwritten character recognition

- size = 100 samples
- choice : attributes = raw pixels
- the size of the images is 32×32
- dimension = 1024 attributes

[image source: P. Geurts, "An introduction to Machine Learning"]

The intrinsic difficulty of machine learning

The theoretical rule to minimize the error rate is

$$y\left(\overrightarrow{x}\right) = \underset{y_i \in \{0,1,\ldots\}}{\arg\max} \left(P\left[y = y_i | \overrightarrow{x}\right]\right) \tag{1}$$

Let ρ be the probability density function of all objects in the attributes space, and ρ_i be the probability density function of the objects belonging to class y_i. Using Bayes' rule,

$$P\left[y = y_i | \overrightarrow{x}\right] = \frac{\rho_i\left[\overrightarrow{x}\right] \; P\left[y = y_i\right]}{\rho\left[\overrightarrow{x}\right]} \tag{2}$$

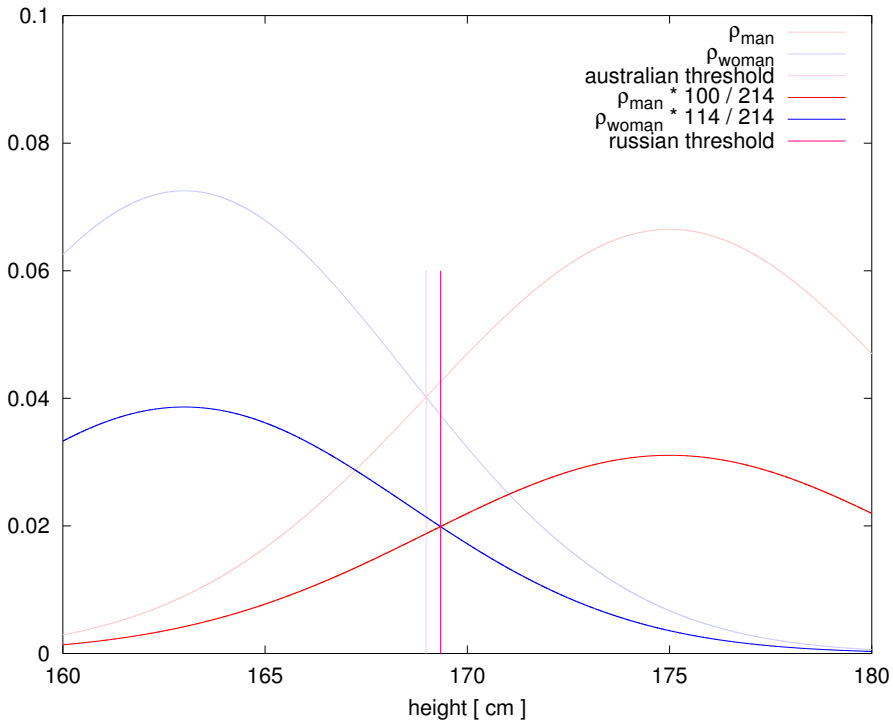
Therefore,

$$y\left(\overrightarrow{x}\right) = \underset{y_i \in \{0,1,\ldots\}}{\arg\max} \left(\rho_i\left[\overrightarrow{x}\right] \; P\left[y = y_i\right]\right) \tag{3}$$

The intrinsic difficulty is that it is very difficult to estimate ρ_i from the learning database because the space is not densely sampled.

An example of decision rule in 1D

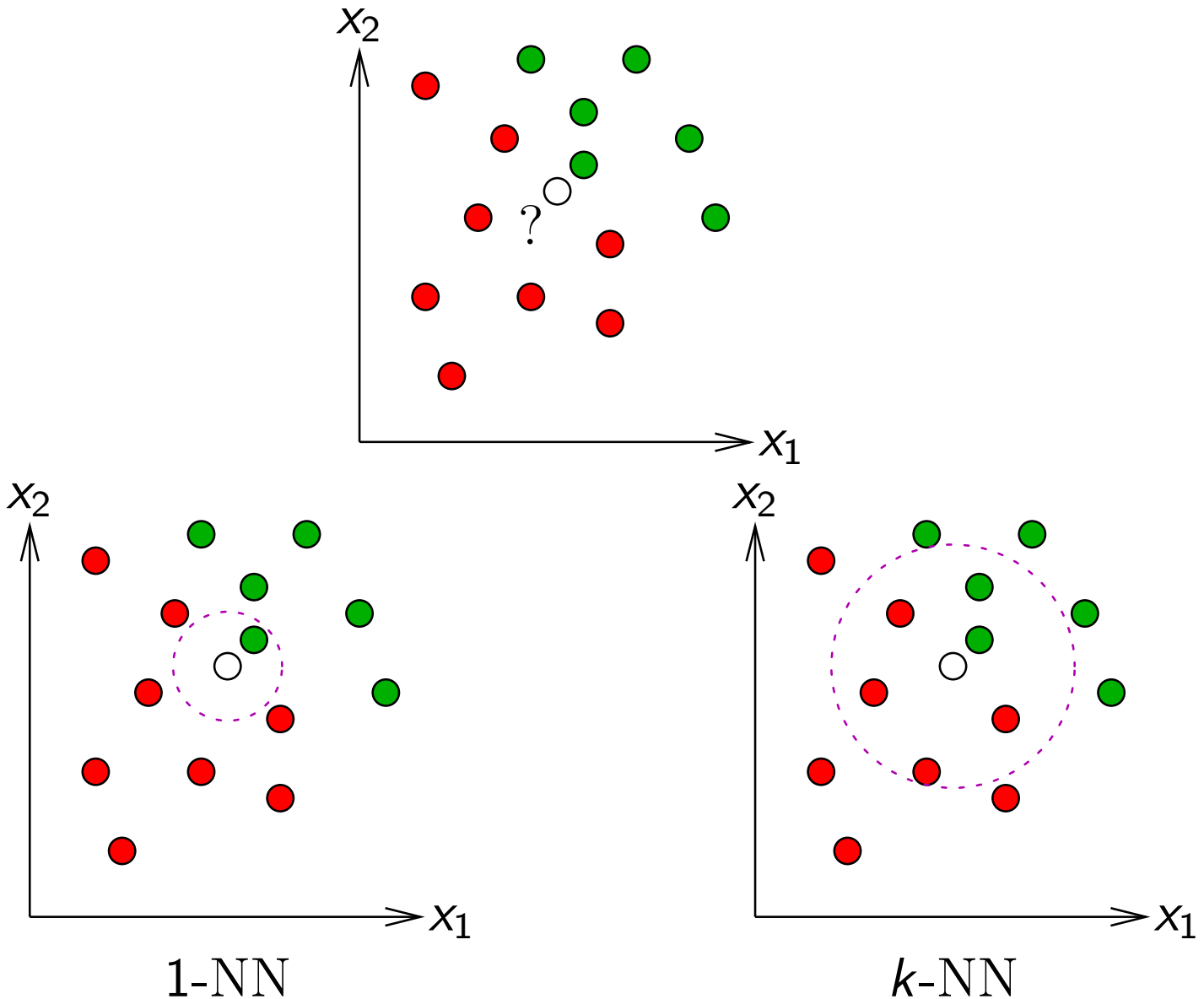
Imagine we have to recognize men and women based on a single attribute: the height. In Australia, there are 100 women for 100 men. But in Russia, there are 114 women for 100 men.



$$y = (\ \text{height} < 169.34\ )\ ?\ \text{"woman"} : \text{"man"}\ ;$$

Example of classifier : the nearest neighbor(s)

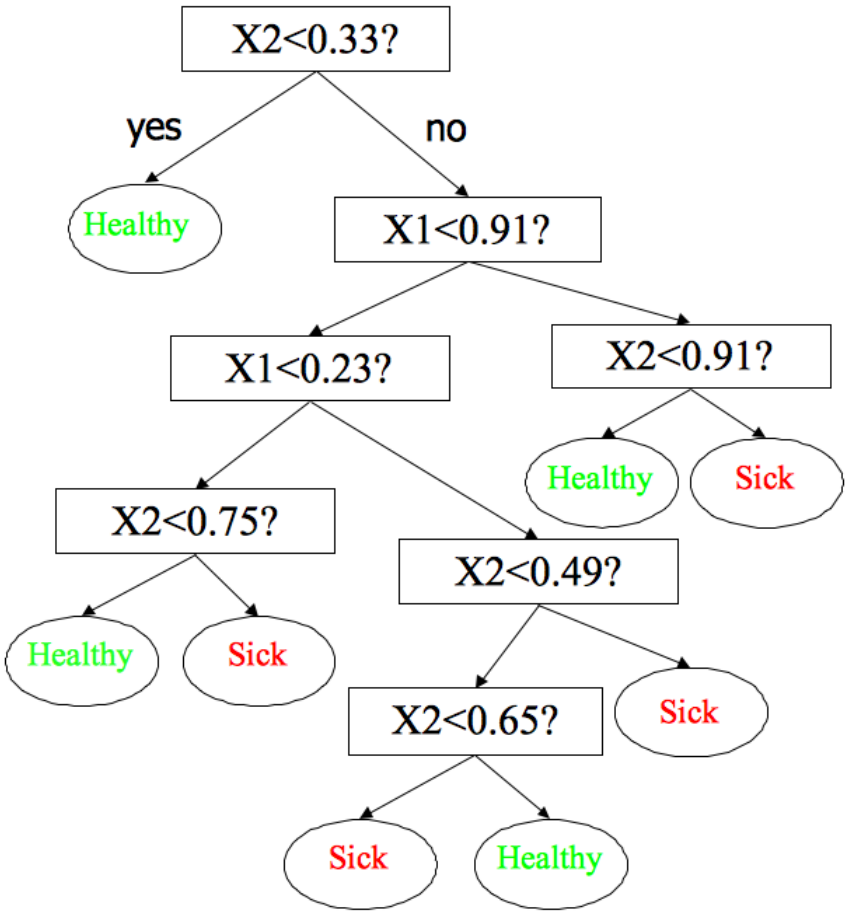
Lets us consider a problem in 2 dimensions:

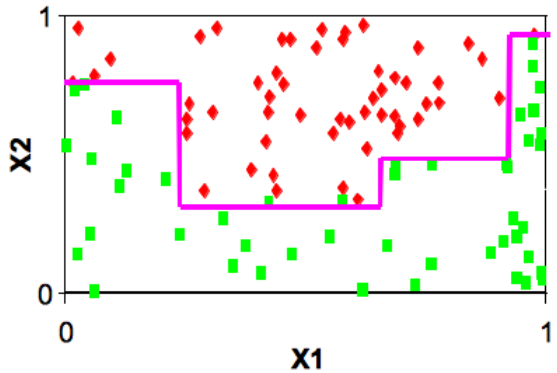


Example of classifier : the nearest neighbor(s)

- The size of the neighborhood is automatically chosen depending on k.
- The model is the learning set (or a pruned version of it).
- First drawback: the time needed to take a decision is $\mathcal{O}(n)$, where n is the learning set size.
- Second drawback: which distance measure should we select ? There exists an infinity of possible choices ! [4]

Example of classifier : the decision trees

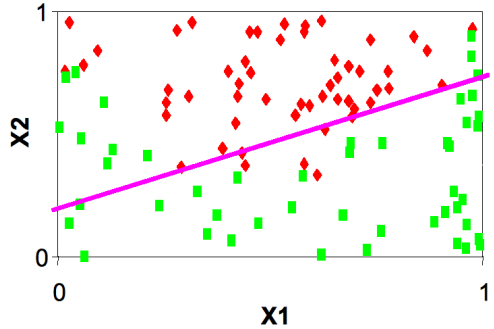


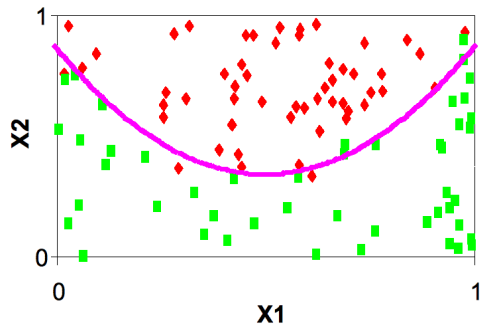


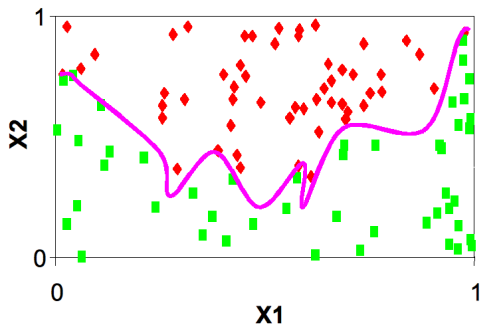
[image source: P. Geurts, "An introduction to Machine Learning"]

Choosing the complexity of the model

Which model is the best ?







[image source: P. Geurts, "An introduction to Machine Learning"]

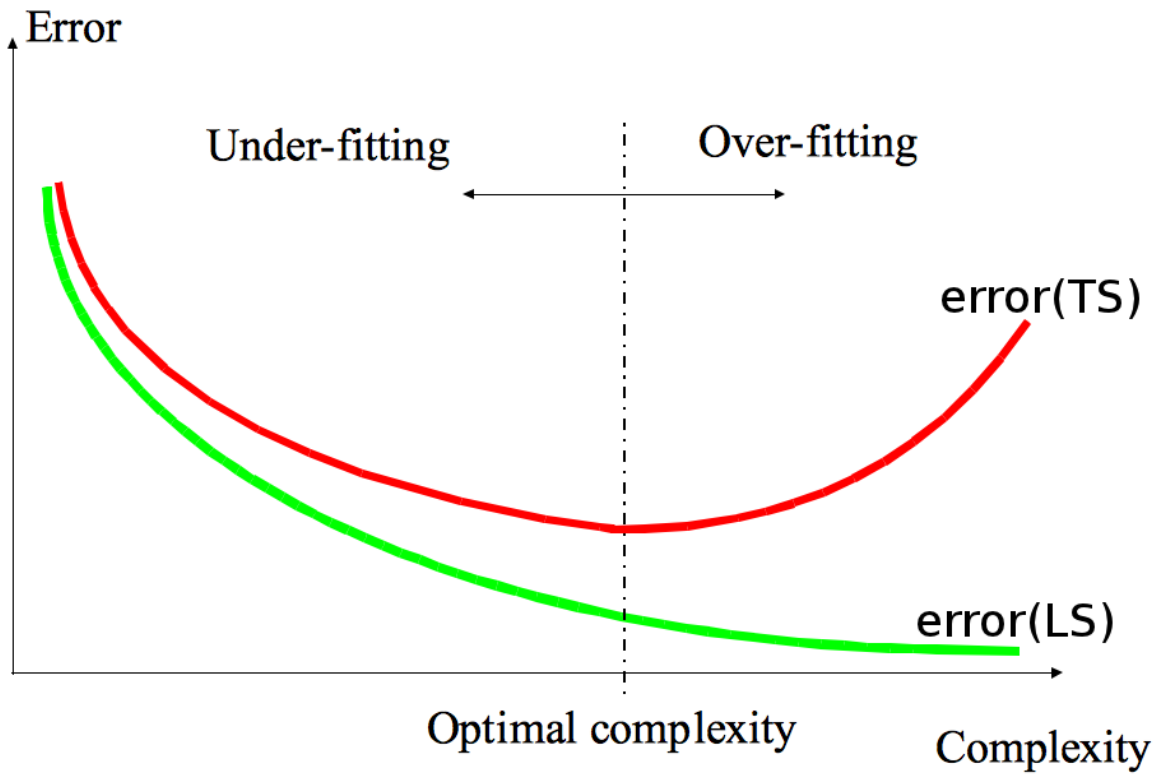
$error(LS) = 3.4\,\%$ $error(TS) = 3.5\,\%$

$error(LS) = 1.0\,\%$ $error(TS) = 1.5\,\%$

$error(LS) = 0.0\,\%$ $error(TS) = 3.5\,\%$

- Does the model explain the learning set?
  $\rightarrow$ *resubstitution error* = error estimated on the learning set
- Is the model able to predict the classes for unknown samples?
  $\rightarrow$ *generalization error* = error estimated on the test set

Choosing the complexity of the model



[image source: P. Geurts, "An introduction to Machine Learning"]

Evaluation of binary classifiers

- The two classes are { positive , negative }
- Example: P = human, N = non-human

		classified as	
		positive	negative
real class (ground truth)	positive	true positives (TP)	false negatives (FN)
	negative	false positives (FP)	true negatives (TN)

$$\#P = \#TP + \#FN \qquad \#N = \#TN + \#FP$$

$$TPR = \frac{\#TP}{\#P} \qquad FNR = \frac{\#FN}{\#P} \qquad TNR = \frac{\#TN}{\#N} \qquad FPR = \frac{\#FP}{\#N}$$

Evaluation of binary classifiers

Remark 1 :

To evaluate a classifier, two quantities are required:

1. TPR or FNR ($TPR + FNR = 1$)
2. TNR or FPR ($TNR + FPR = 1$)

Remark 2 :

There is always a trade-off :

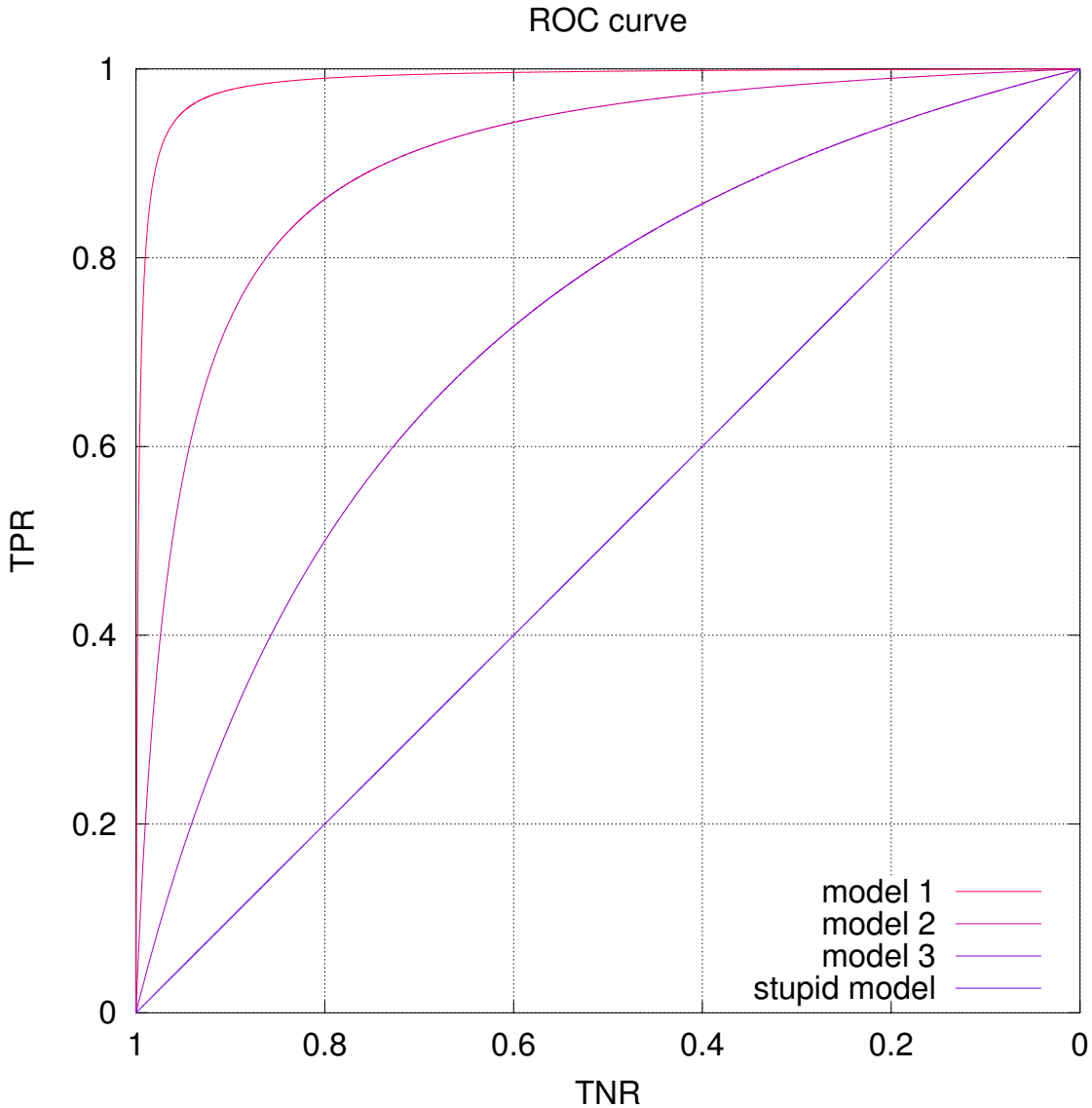
- It is easy to obtain a high TPR.
- It is easy to obtain a high TNR.
- But it is difficult to obtain both simultaneously !

Remark 3 :

A binary classification = a threshold thr. Both TPR and TNR depend on the value of thr. Therefore we need to carefully choose the value of thr to optimize (TPR, TNR) !

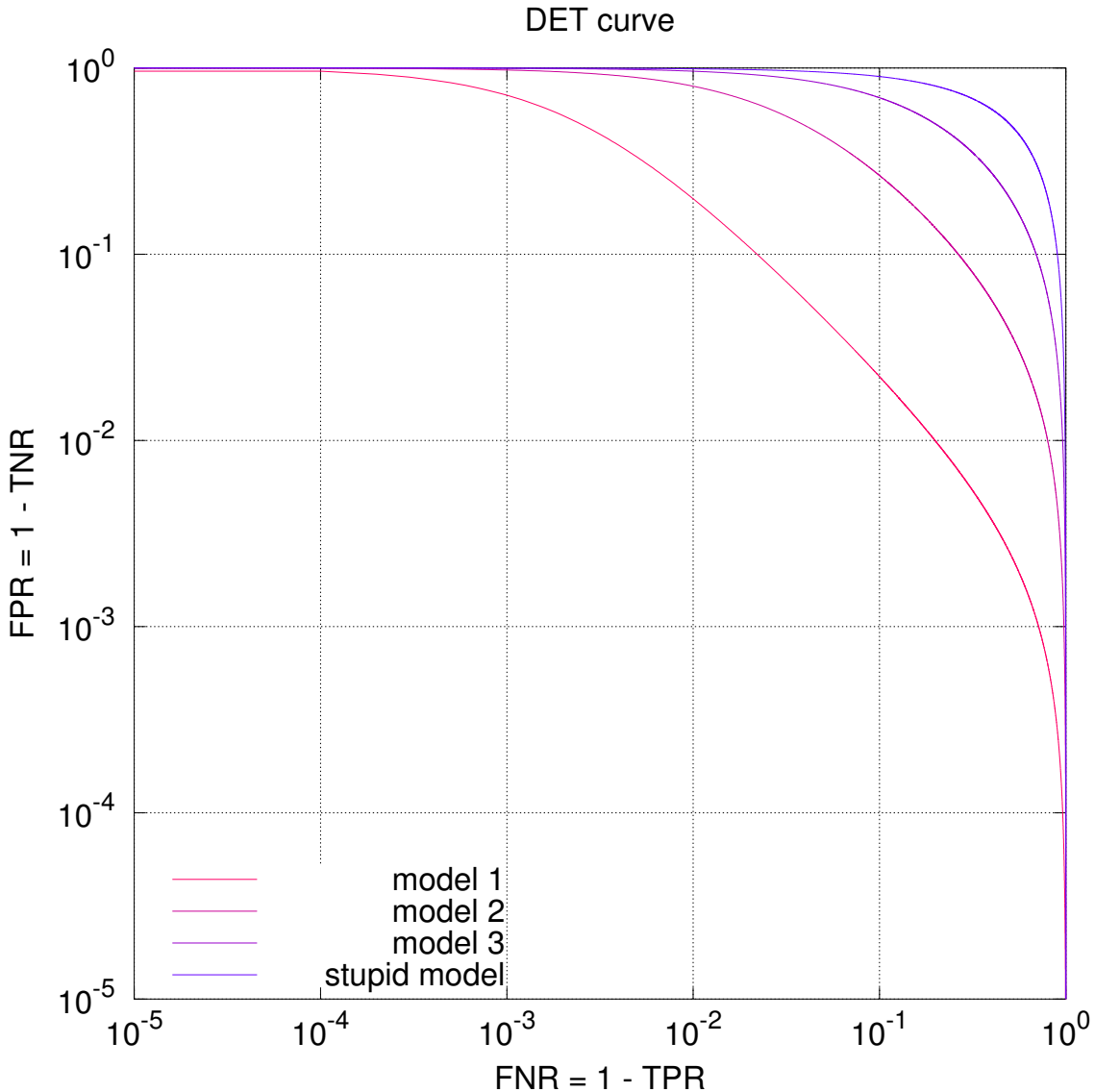
Evaluation: receiver operating characteristic (ROC)

$$\{(TPR(thr), TNR(thr)) \forall thr \in \mathbb{R}\} = \text{a ROC curve}$$



Evaluation: detection error tradeoff (DET)

$$\{(TPR(thr), TNR(thr)) \forall thr \in \mathbb{R}\} = \text{a DET curve}$$



Outline

Conclusion

ML = automatic + generalization + preprocessing

Machine learning techniques are :

- powerful methods;
- a complement to traditional algorithmics;
- indispensable in computer science;
- adequate for real-time computations;
- "easy" to use[2].

[2] However, optimal results are difficult to obtain. This is why researchers are still working on machine learning methods.

Bibliography I


N. Boulgouris, D. Hatzinakos, and K. Plataniotis.
Gait recognition: a challenging signal processing technology for biometric identification.
IEEE Signal Processing Magazine, 22(6):78–90, November 2005.


C. Cortes and V. Vapnik.
Support-vector networks.
Machine Learning, 20(3):273–297, 1995.


N. Dalal and B. Triggs.
Histograms of oriented gradients for human detection.
In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, San Diego, USA, June 2005.

Bibliography II


M. Deza and E. Deza.
Encyclopedia of Distances.
Springer, 2009.


P. Geurts, D. Ernst, and L. Wehenkel.
Extremely randomized trees.
Machine Learning, 63(1):3–42, April 2006.


T. Hastie, R. Tibshirani, and J. Friedman.
The elements of statistical learning: data mining, inference, and prediction.
Springer Series in Statistics. Springer, second edition, September 2009.

Bibliography III

J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake.
Real-time human pose recognition in parts from single depth images.
In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, June 2011.

P. Viola and M. Jones.
Robust real-time face detection.
International Journal of Computer Vision, 57(2):137–154, 2004.