introduction to Machine Learning

What

3 lectures:

- 1: general introduction
- 2: bayesian methods and parametric gaussian models
- 3: non parametric models

Who

Ms students and last year engineers school

Credits

A. Zisserman lecture (oxford)

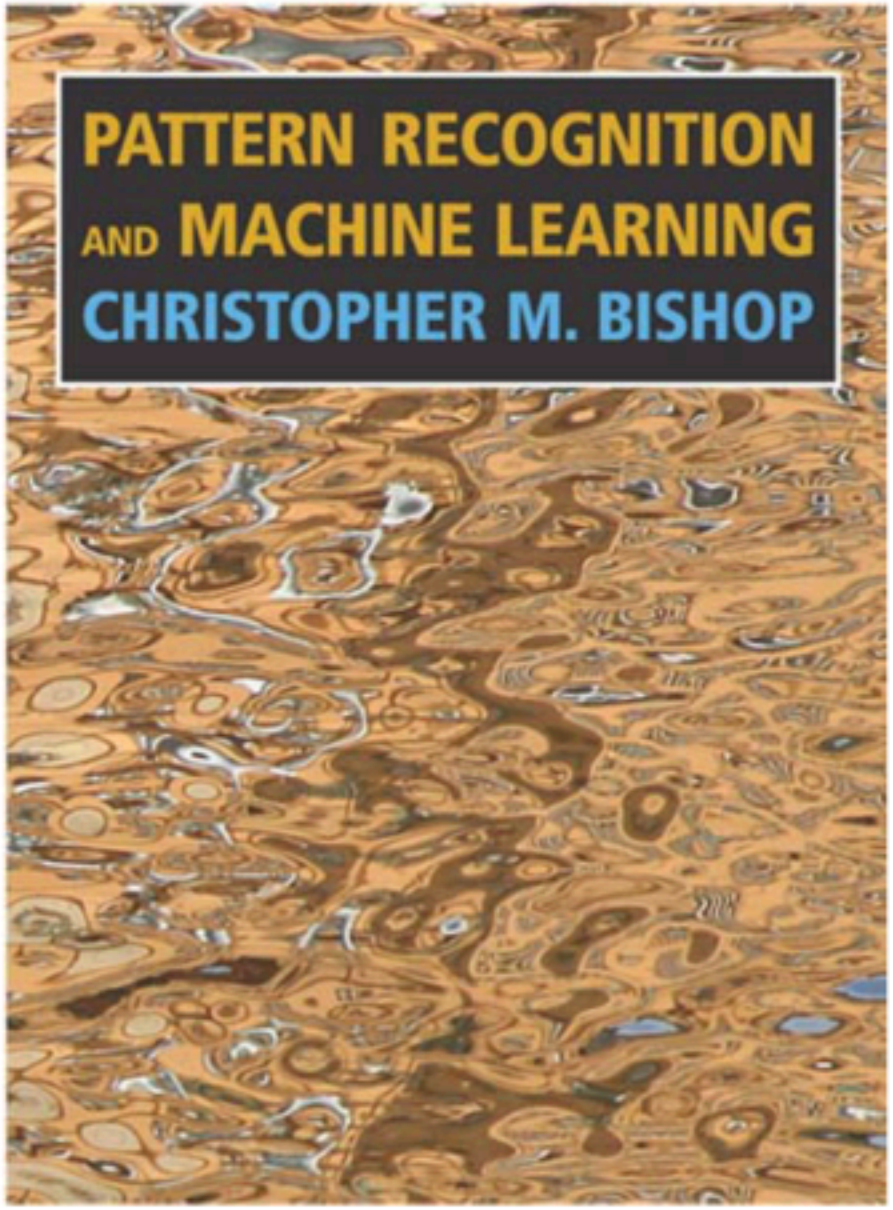
C. Wolf: LIRIS, Lyon

recommended books

- **Pattern Recognition and Machine Learning**

Christopher Bishop, Springer, 2006.

  - Excellent on classification and regression



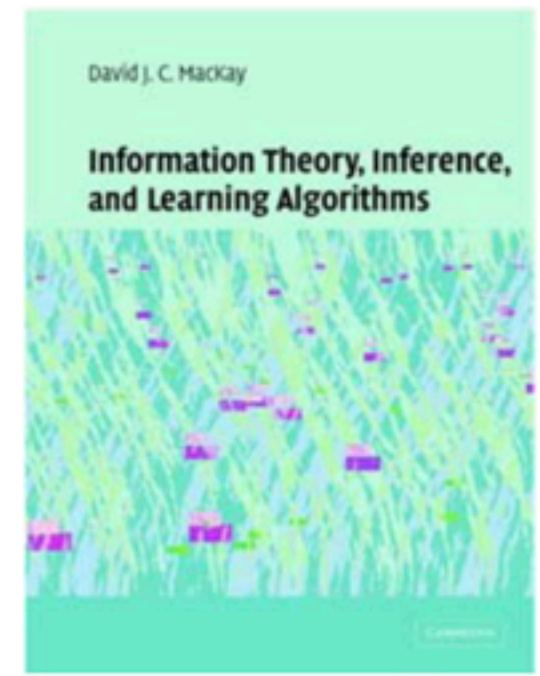
recommended books

- On line book:

Information Theory, Inference, and Learning Algorithms.

David J. C. MacKay, CUP, 2003

  - Covers some of the course material though at an advanced level



What is Machine Learning?

an algorithm that can improve it performance using training datas

$$\mathbf{x} \rightarrow \boxed{\mathbf{y} = f(\mathbf{x};\boldsymbol{\theta})} \rightarrow \mathbf{y}$$

$\boldsymbol{\theta}$ is a vector of parameters (large) computed from a training database

function f cannot be defined with *rules by hands*

face detection, speech recognition, stock prediction,…

What is Machine Learning?

an algorithm that can improve it performance using training datas

$$\mathbf{x} \rightarrow \mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta}) \rightarrow \mathbf{y}$$

$\boldsymbol{\theta}$ **is a vector of parameters (large) computed from a training database**

if y is a discret: classification
if y is continus: regression

Machine Learning

The machine learning framework

- Apply a prediction function to a feature representation of the image to get the desired output:

f() = “apple”

f() = “tomato”

f() = “cow”

Slide credit: D. Hoiem and L. Lazebnik

Slide credit: L. Lazebnik

Traditional Machine Learning

Training

Training Images

Image Features → Training → Learned model

Training Labels → Training

Testing

Test Image

Image Features → Learned model → Prediction

Slide credit: D. Hoiem and L. Lazebnik

Recent Machine Learning (IA)

Training

Training Images

Training Labels

Training

Learned model

Testing

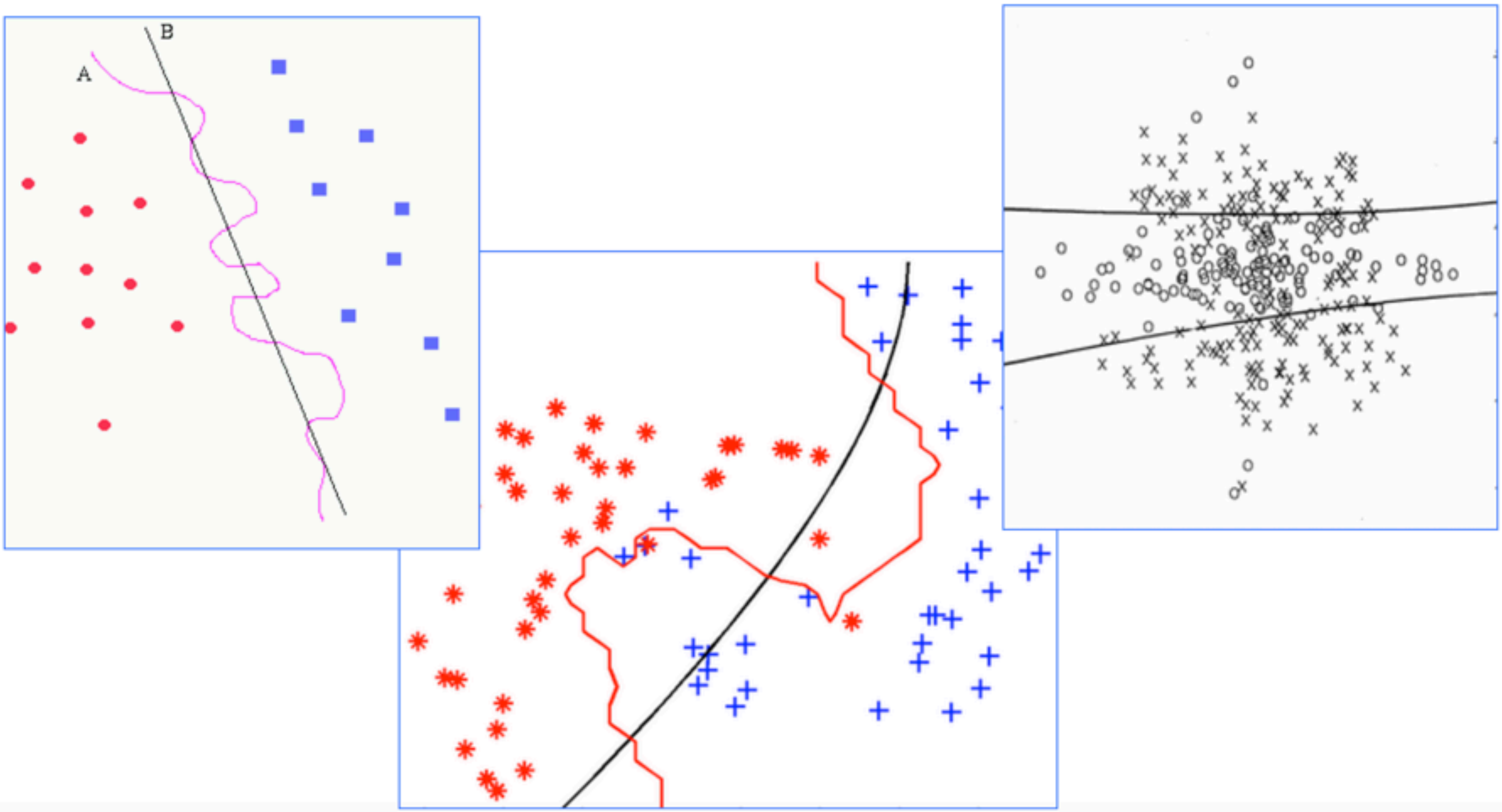
Test Image

Learned model

Prediction

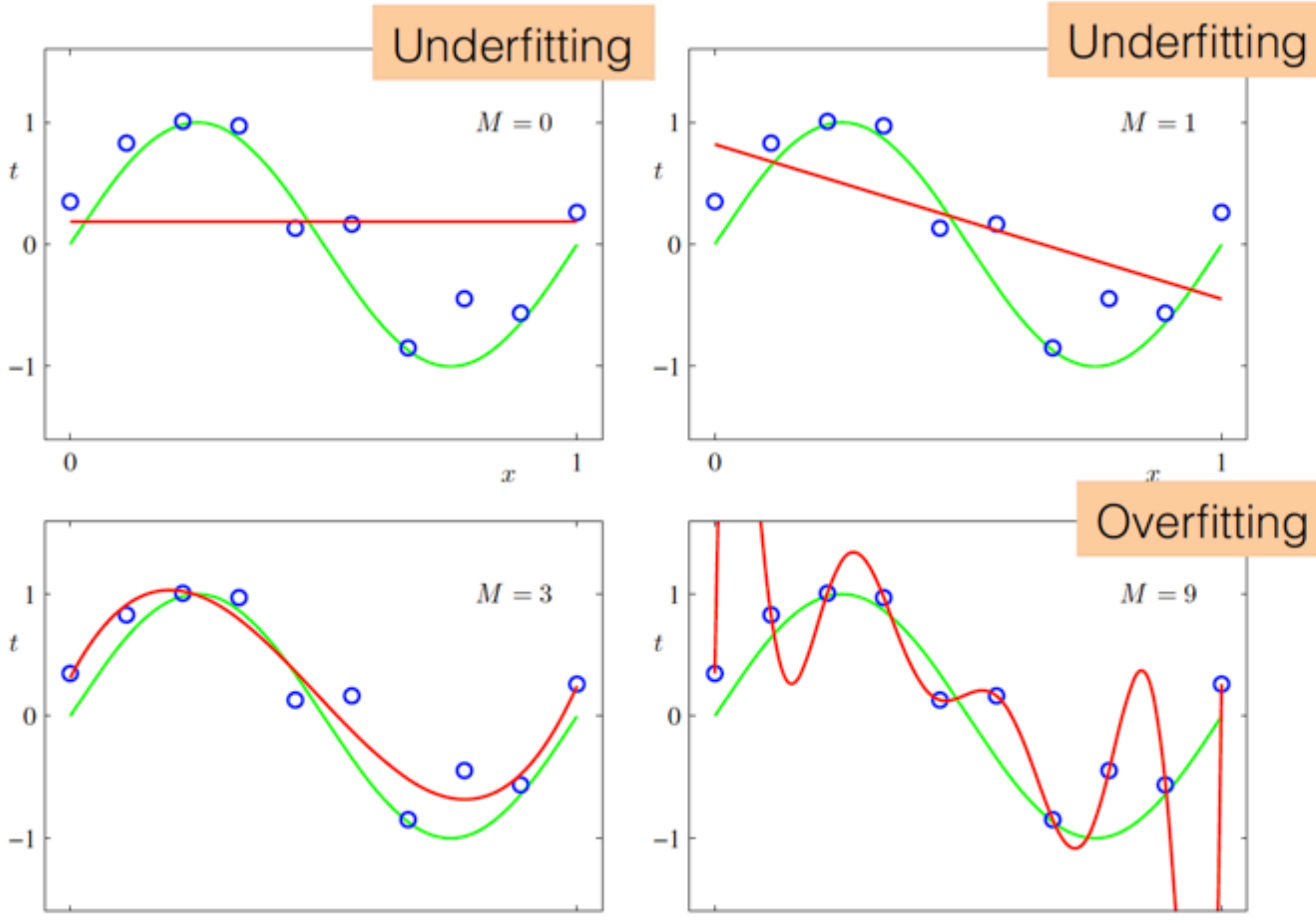
Slide credit: D. Hoiem and L. Lazebnik

Learn a decision function into a feature space



slide credits: C. Wolf

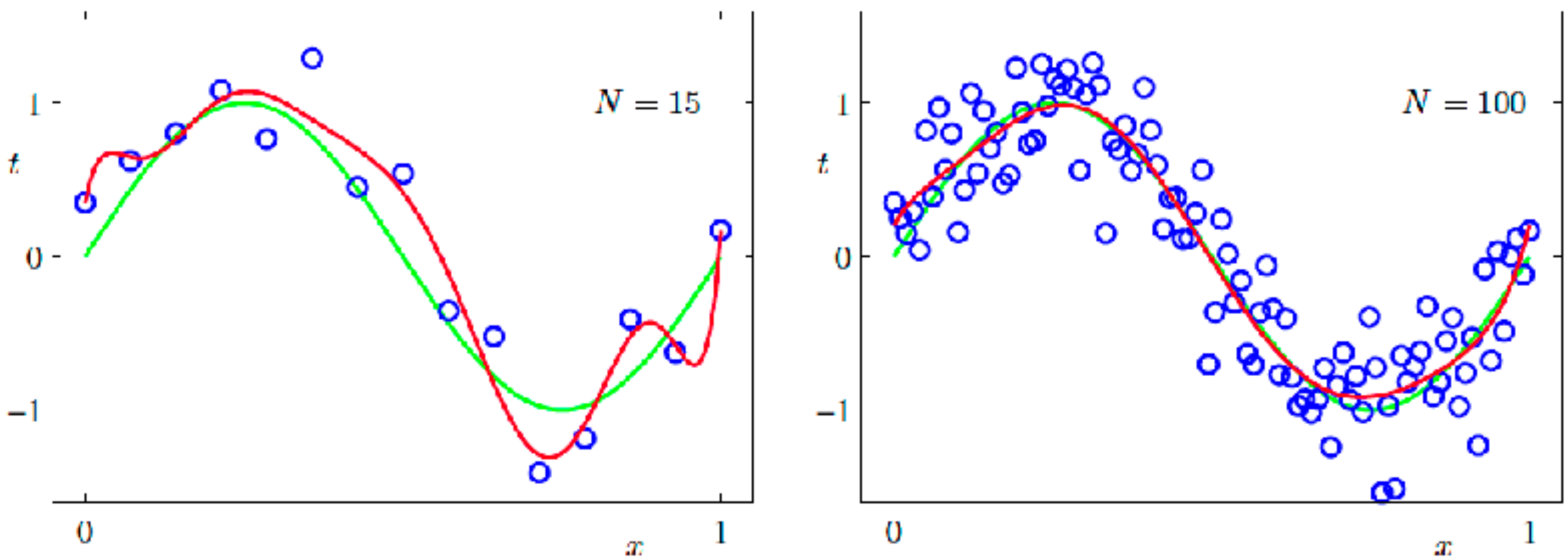
Model selection



$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

slide credits: C. Wolf

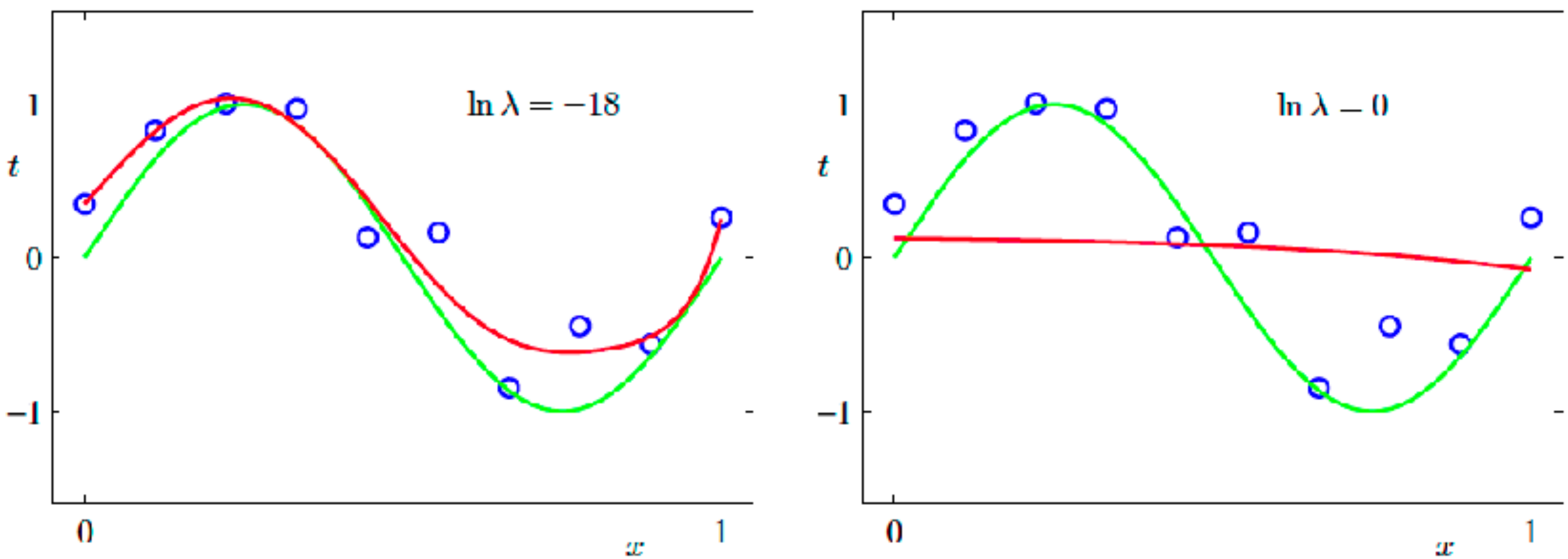
Overfitting can be reduced by increasing the training size



slide credits: C. Wolf

Regularisation

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2$$

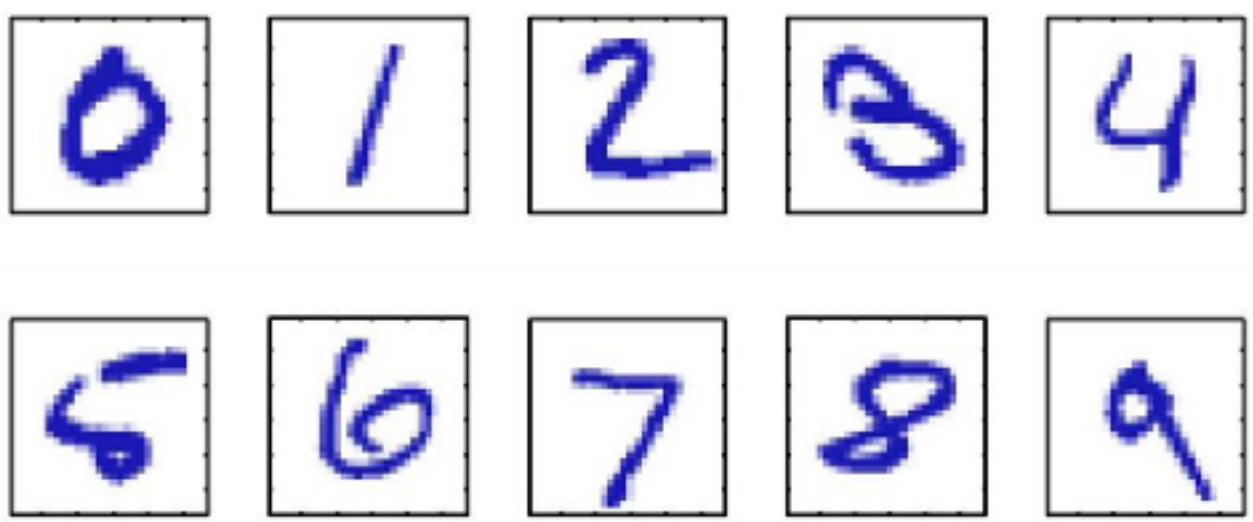


slide credits: C. Wolf

Introduction

Example 1:
hand written digit recognition

$$\mathbf{x} \longrightarrow \boxed{\mathbf{y} = f(\mathbf{x};\boldsymbol{\theta})} \longrightarrow \mathbf{y}$$



Images are 28 x 28 pixels

represent input image as a vector:

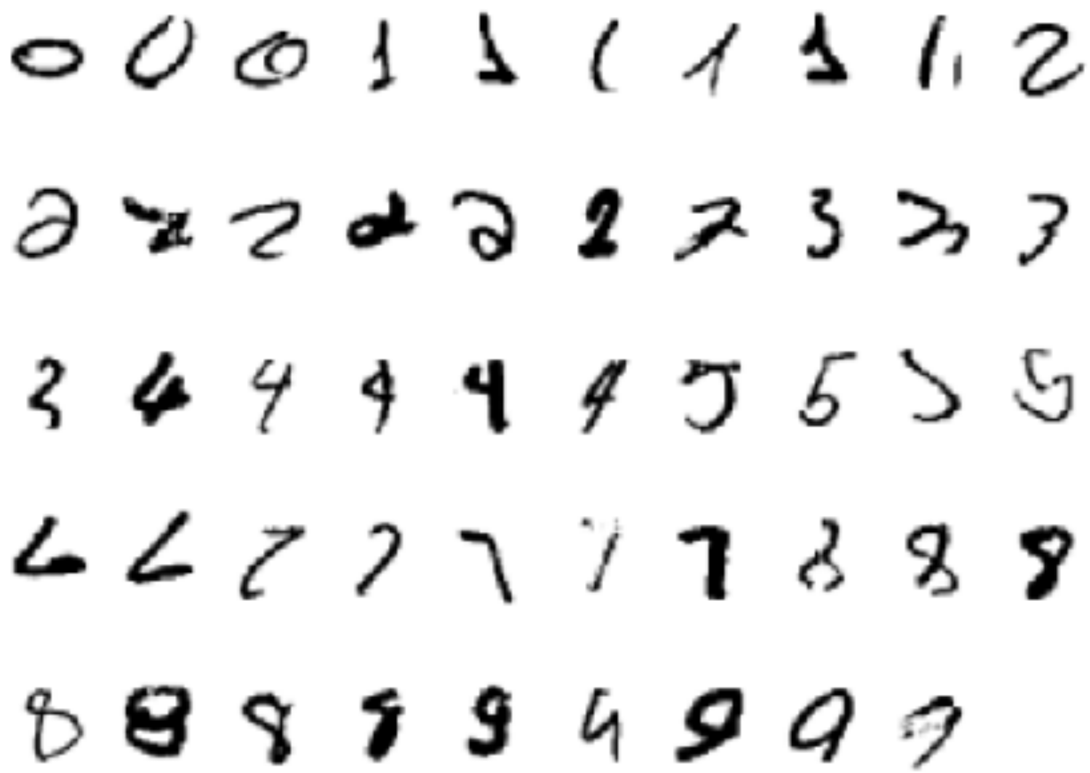
$$\mathbf{x} \in I\!R^{794}$$

learn function f:

$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

this is a **classification** problem

Example 1:
hand written digit recognition



$$\mathbf{x} \rightarrow \boxed{\mathbf{y} = f(\mathbf{x};\boldsymbol{\theta})} \rightarrow \mathbf{y}$$

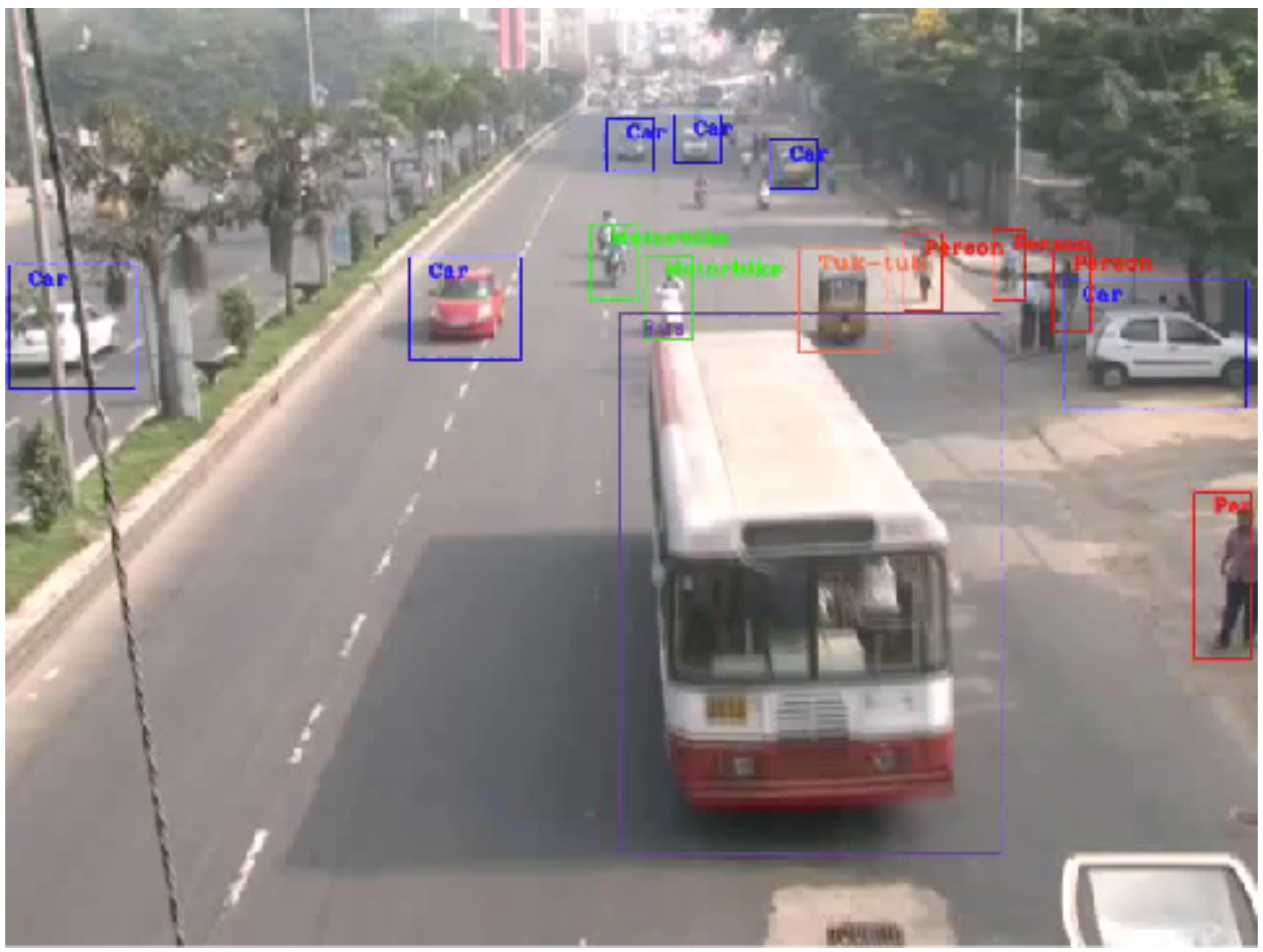
we need an annotated dataset (supervised learning)
6000 samples to learn the parameter vector $\boldsymbol{\theta}$

Training based systems can achieved a test error of 0.4%

Example 2:
vehicle detection

$$\mathbf{x} \rightarrow \mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta}) \rightarrow \mathbf{y}$$

we need an annotated dataset (supervised learning) samples to learn the parameter vector $\boldsymbol{\theta}$

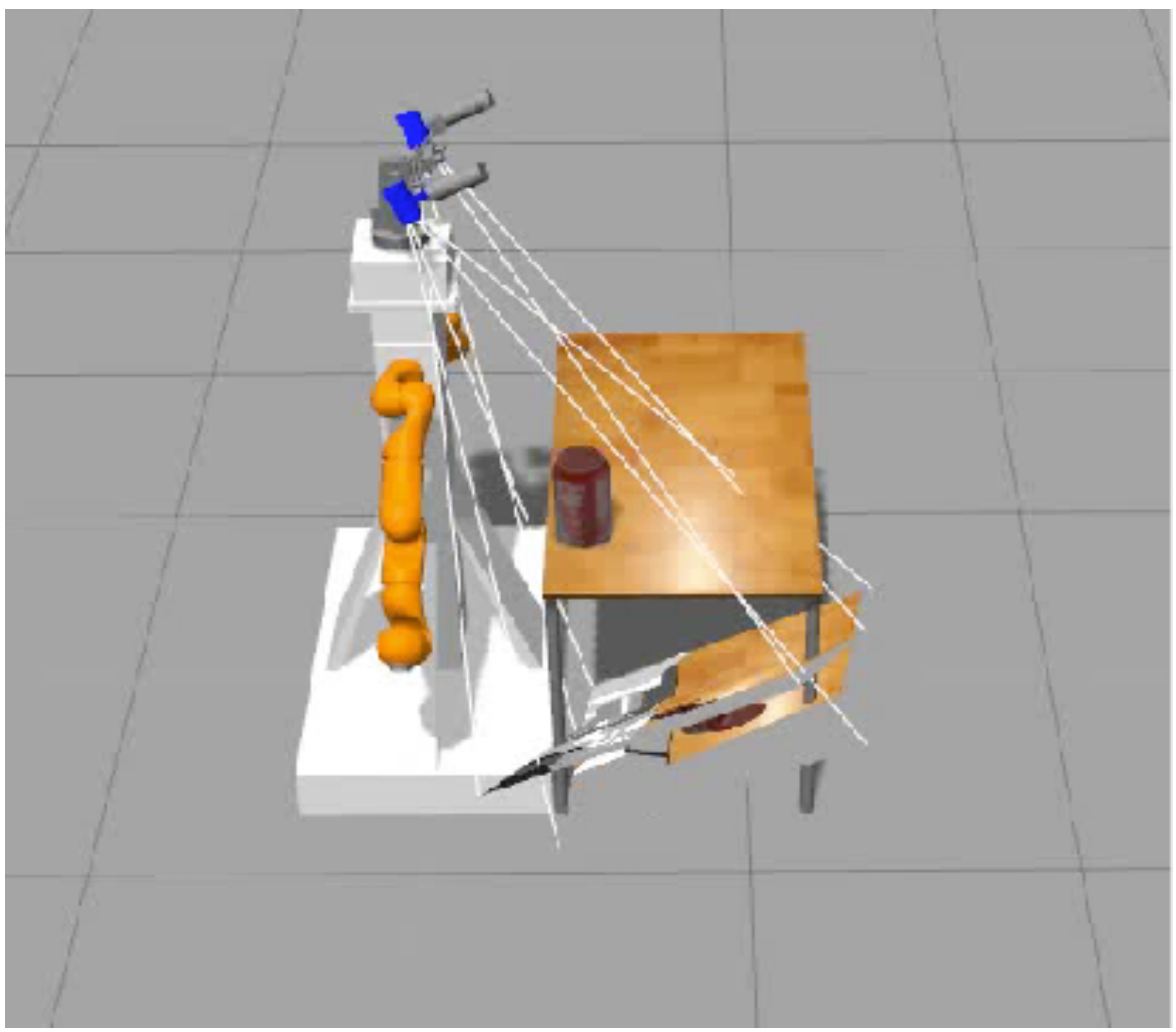


Mhalla PhD: Pascal Institute, 2017

Introduction

Example 3:
sensori-motor estimation

stereo-vision focusing



François de la Bourdonnaye, PhD, 2017; IP

$$\mathbf{x} \rightarrow \boxed{\mathbf{y} = f(\mathbf{x};\boldsymbol{\theta})} \rightarrow \mathbf{y}$$

we need an annotated dataset (supervised learning)
6000 samples to learn the parameter vector $\boldsymbol{\theta}$

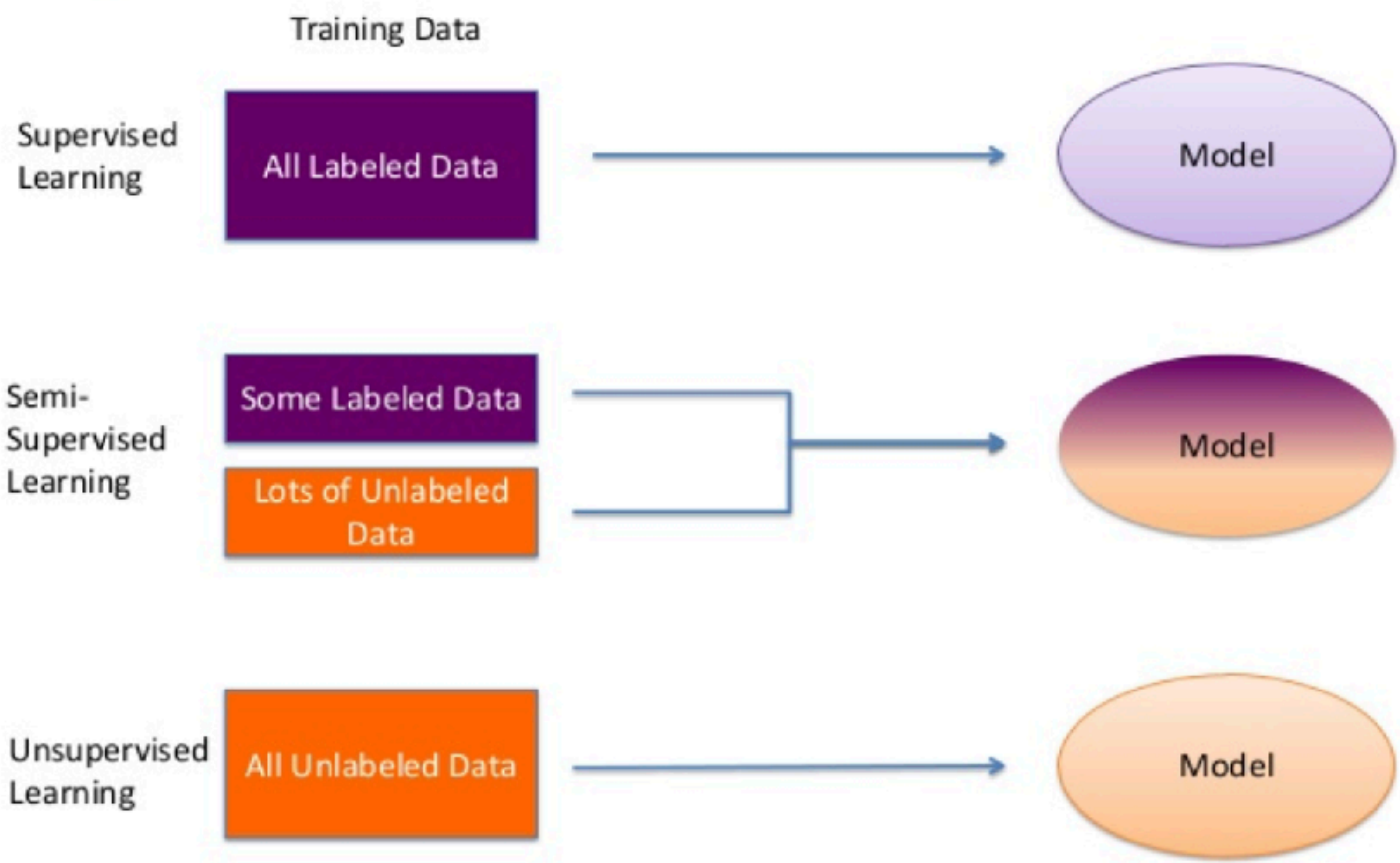
image center
object detection

Timeline of machine learning?

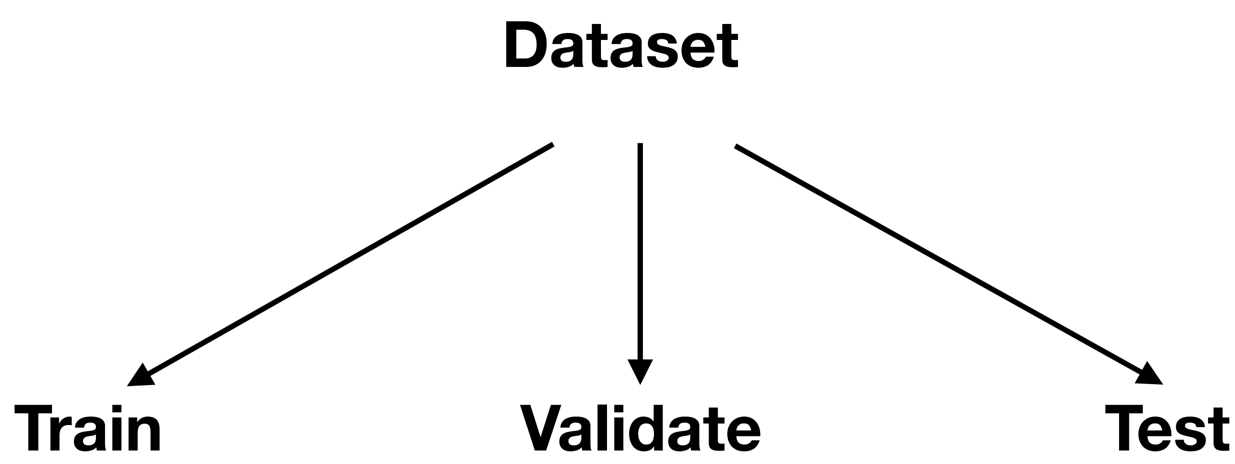
Decade	Summary
<1950s	Statistical methods are discovered and refined.
1950s	Pioneering machine learning research is conducted using simple algorithms.
1960s	Bayesian methods are introduced for probabilistic inference in machine learning[1].
1970s	'AI Winter' caused by pessimism about machine learning effectiveness.
1980s	Rediscovery of backpropagation causes a resurgence in machine learning research.
1990s	Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data Support vector machines and recurrent neural networks become popular.
2000s	Kernel methods grow in popularity[3], and competitive machine learning becomes more widespread[4].
2010s	Deep learning becomes feasible, which leads to machine learning becoming integral to many widely used software services and applications.

https://en.wikipedia.org/wiki/Timeline_of_machine_learning

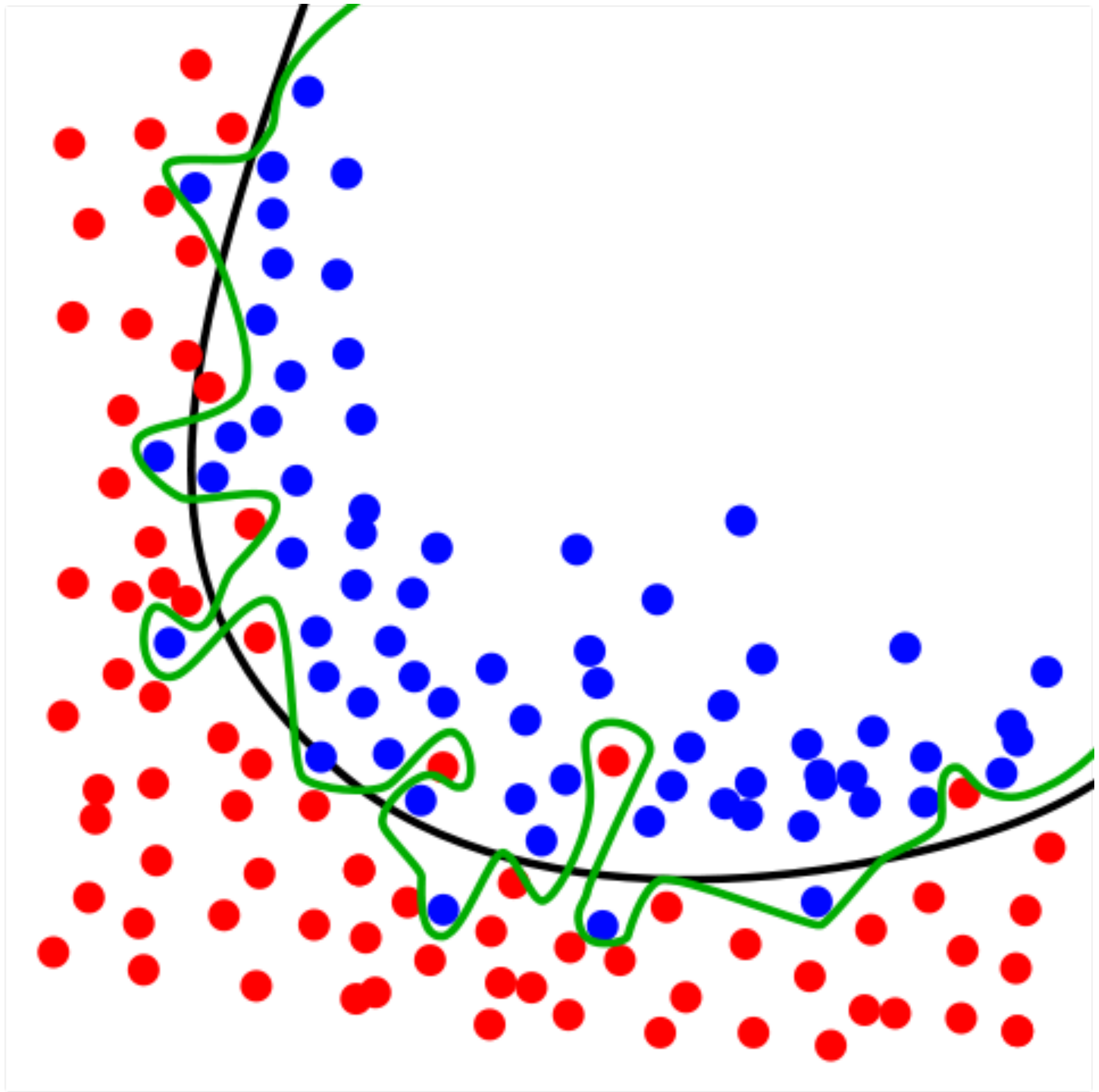
Learning based approaches



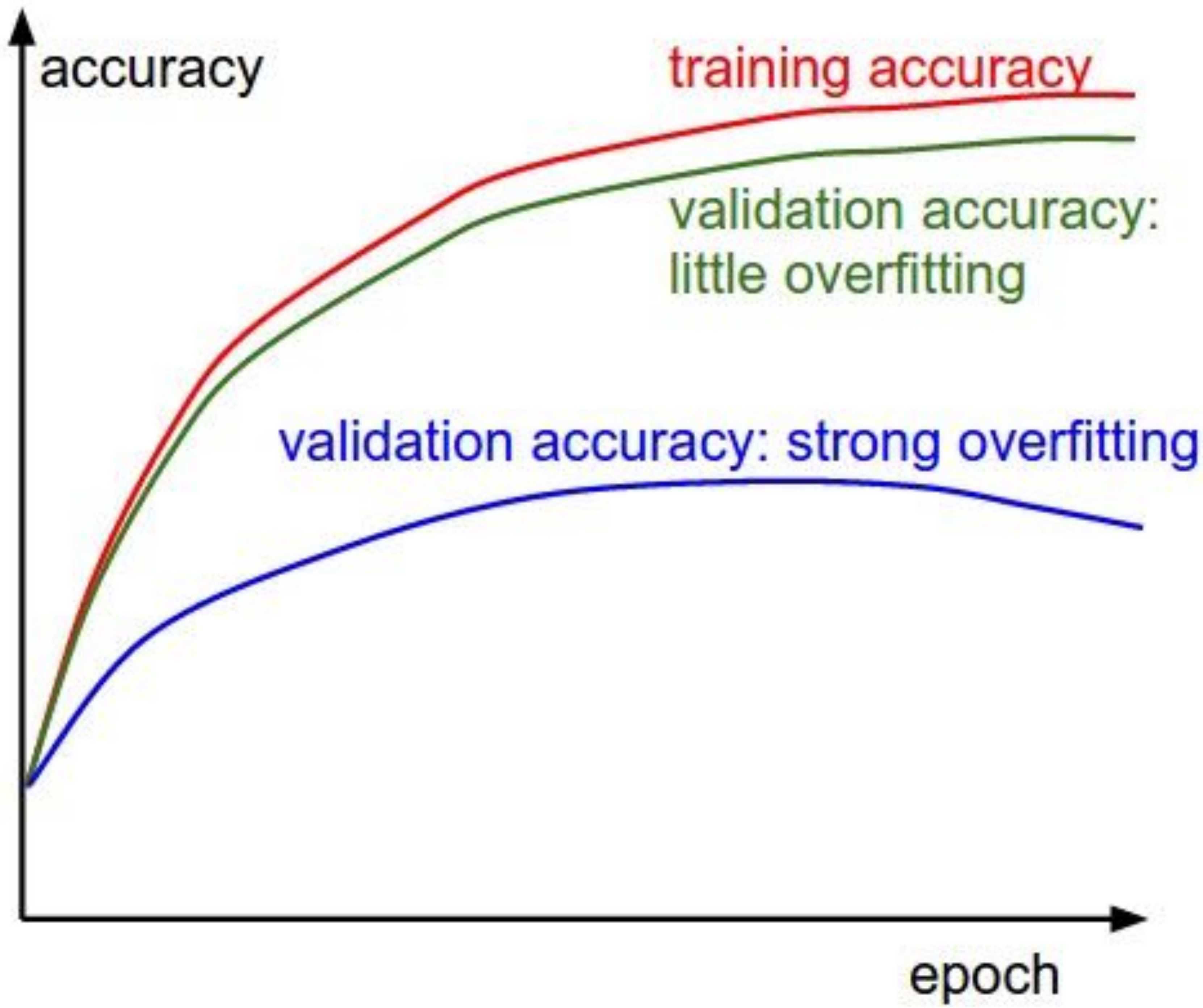
Datasets



Over-learning



Over-learning



Evaluation criteria (classification)

Accuracy on test set:
The rate of correct classification on testing set

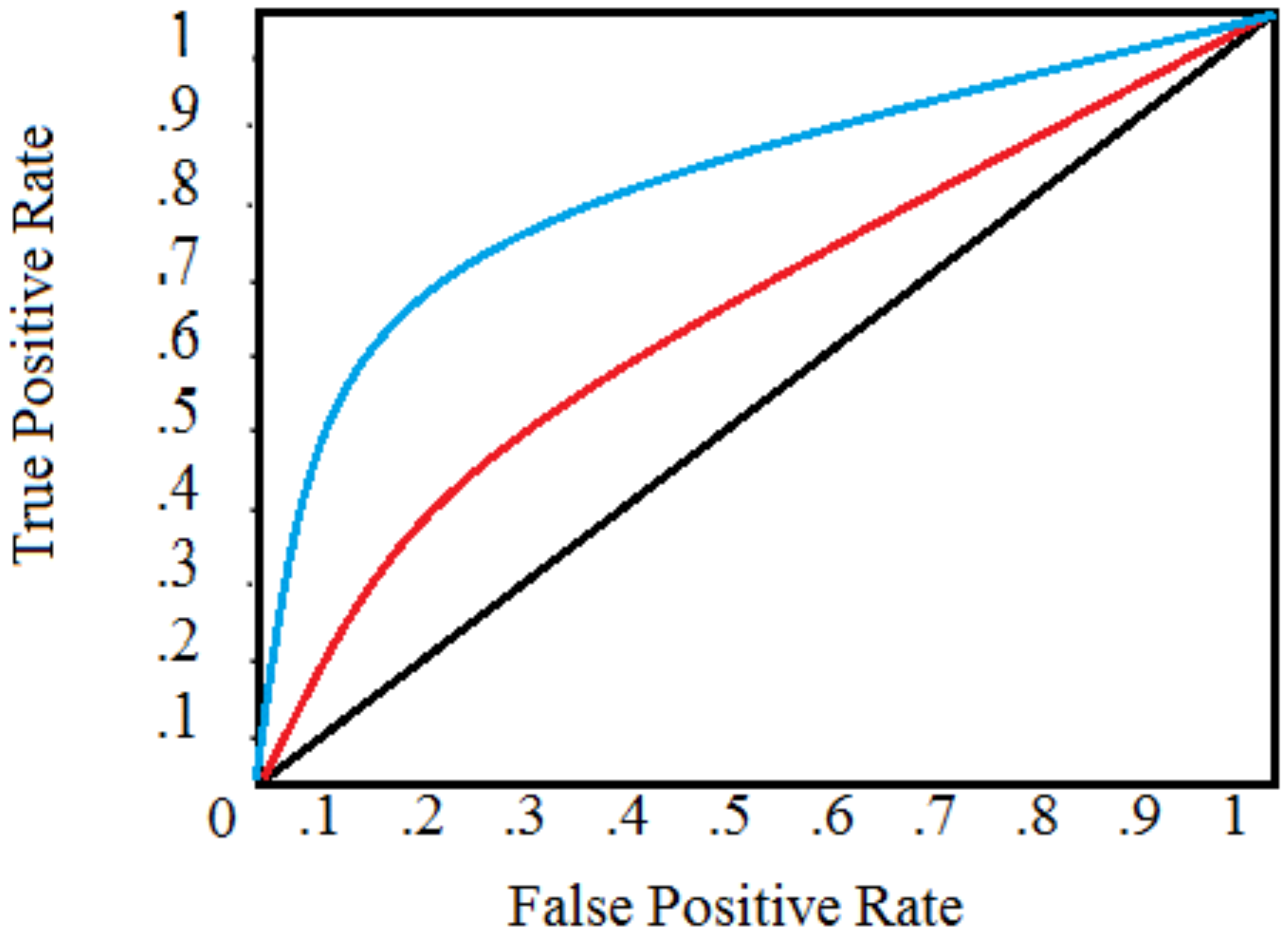
Error Rate on test set:
the percentage of wrong predictions on test set

Confusion matrix

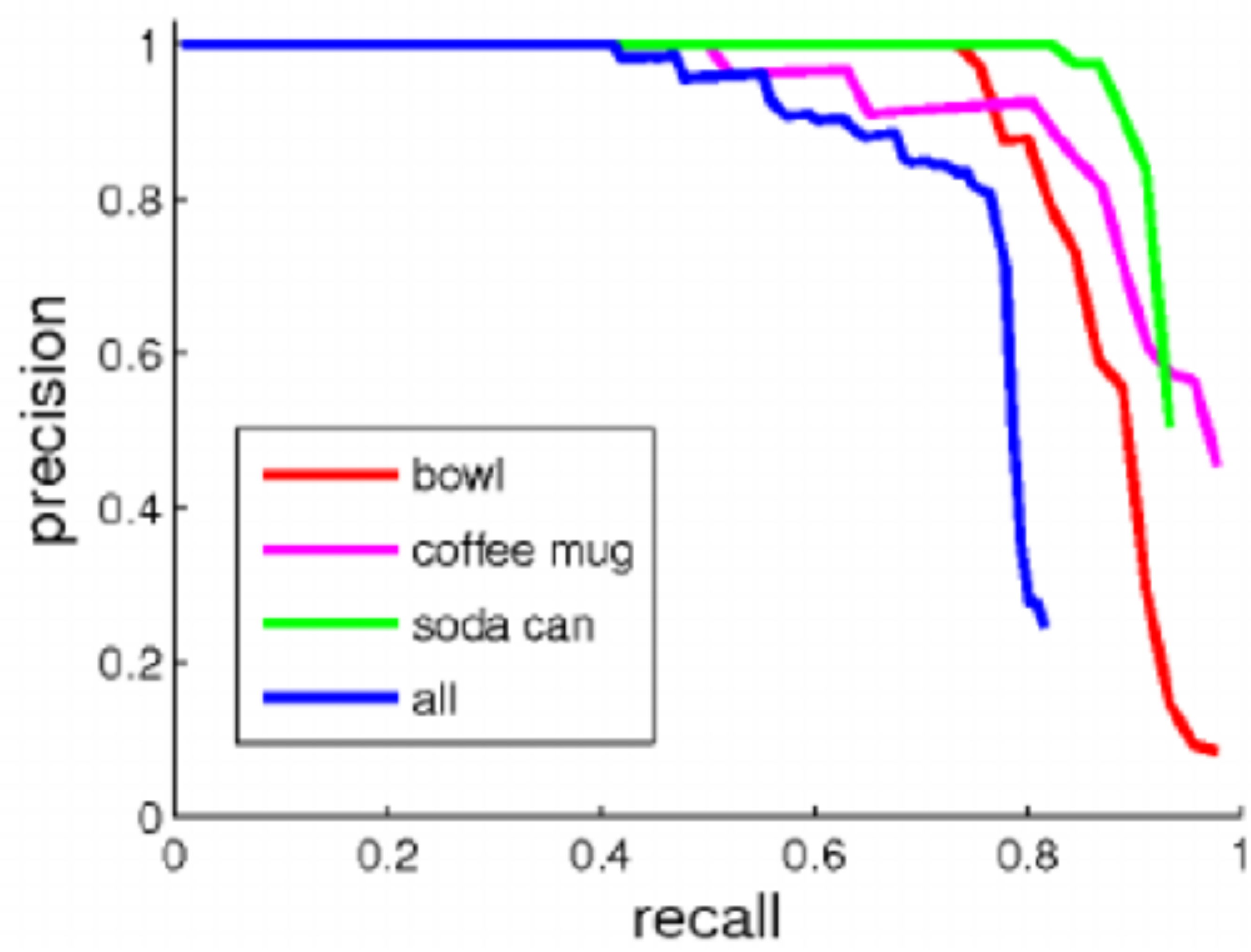
Speed and scalability:
the time to build the classifier and to classify new sample, and the scalability with respect to the data size

Robustness:
handling noise and missing values

Evaluation criteria (ROC curve)



Evaluation criteria (recall precision curve)



$$Precision = \frac{tp}{tp+fp} \text{ and}$$

$$Recall = \frac{tp}{tp+fn}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Evaluation criteria (detection)

Intersection over Union (IoU) for object detection

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Supervised classification

parametric methods
- Bayesian classifiers
- SVM
- Random forest
- Neural networks

non parametric methods
- K nearest neighbours
- Kernel density estimation

Starred algorthims

- bayes rule
- kppv
- svm
- adaboost
- …
- neural networks