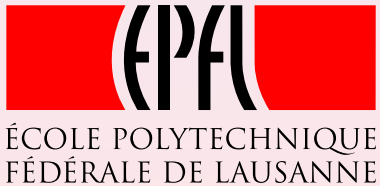
Statistical Machine Learning from Data

Introduction to Machine Learning

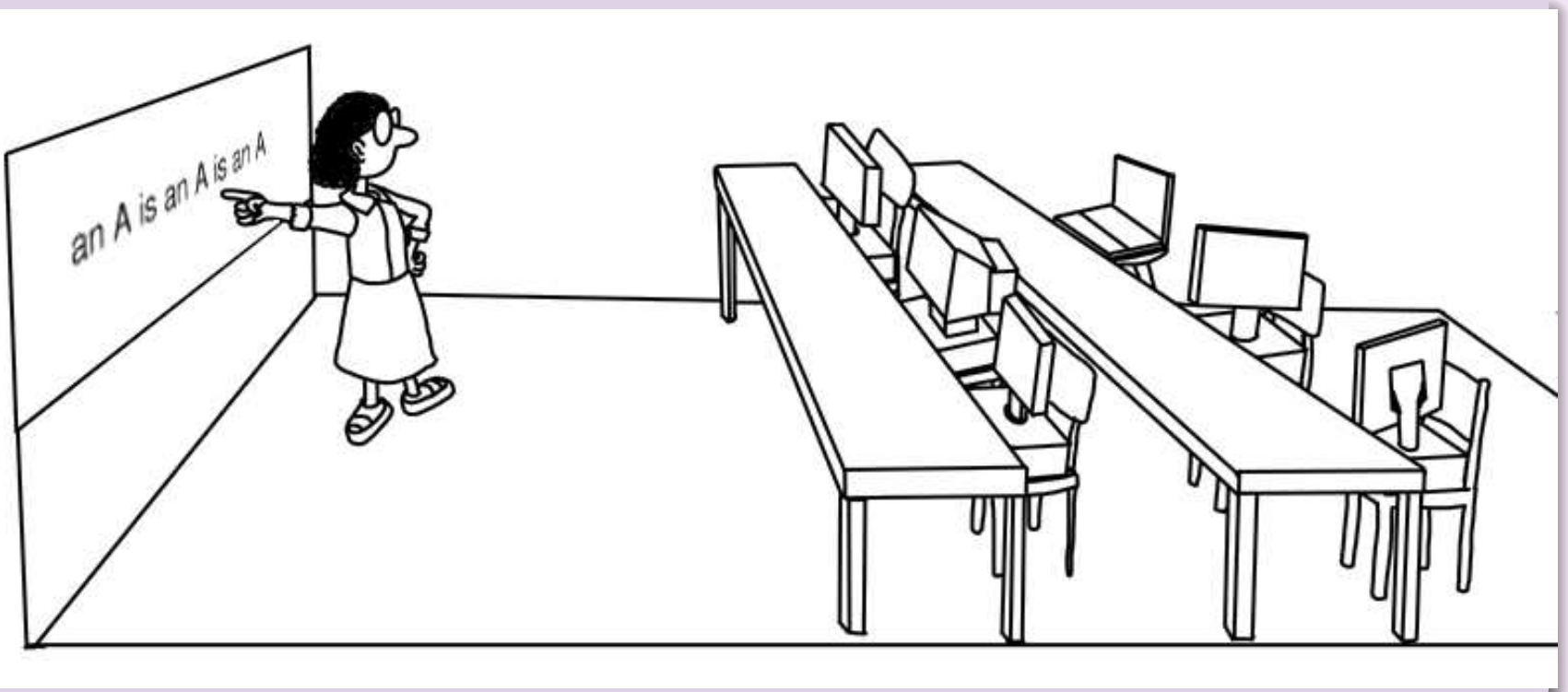
Samy Bengio

IDIAP Research Institute, Martigny, Switzerland, and
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
bengio@idiap.ch
http://www.idiap.ch/~bengio



November 30, 2005

What is Machine Learning? (Graphical View)

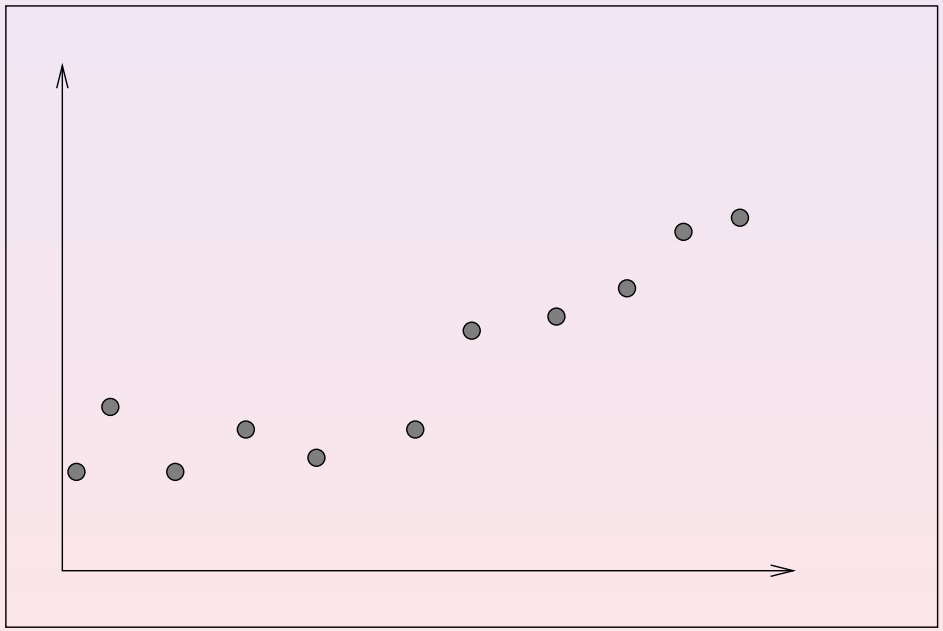


What is Machine Learning?

- Learning is an essential human property
- Learning means changing in order to be better (according to a given criterion) when a similar situation arrives
- Learning IS NOT learning by heart
- Any computer can learn by heart, the difficulty is to generalize a behavior to a novel situation

Why Learning is Difficult?

- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations



- How should we draw the relation?

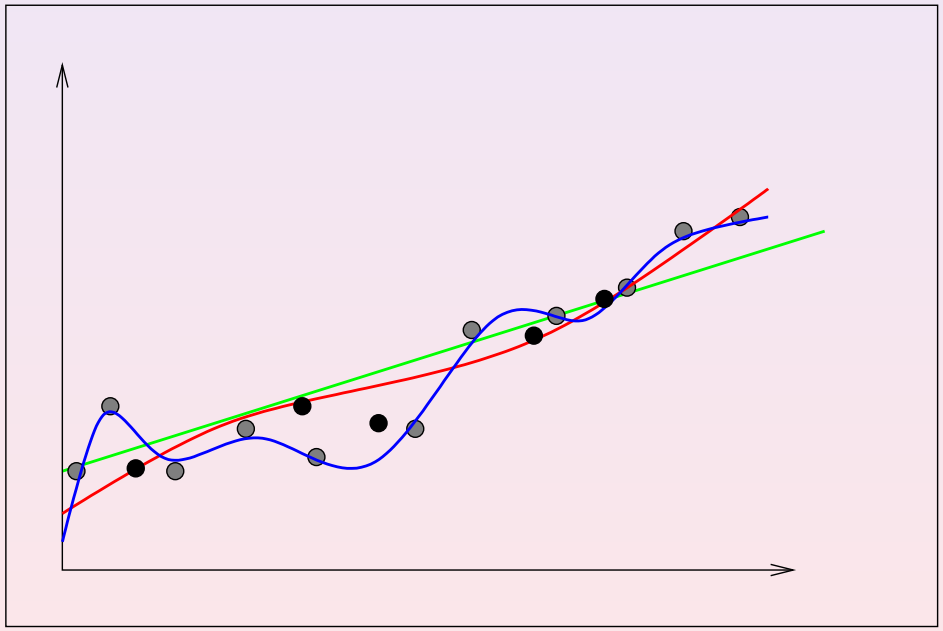
Why Learning is Difficult? (2)

- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations

- Which relation is the most appropriate?

Why Learning is Difficult? (3)

- Given a finite amount of training data, you have to derive a relation for an infinite domain
- In fact, there is an infinite number of such relations



- ... the hidden test points...

Occam's Razor's Principle

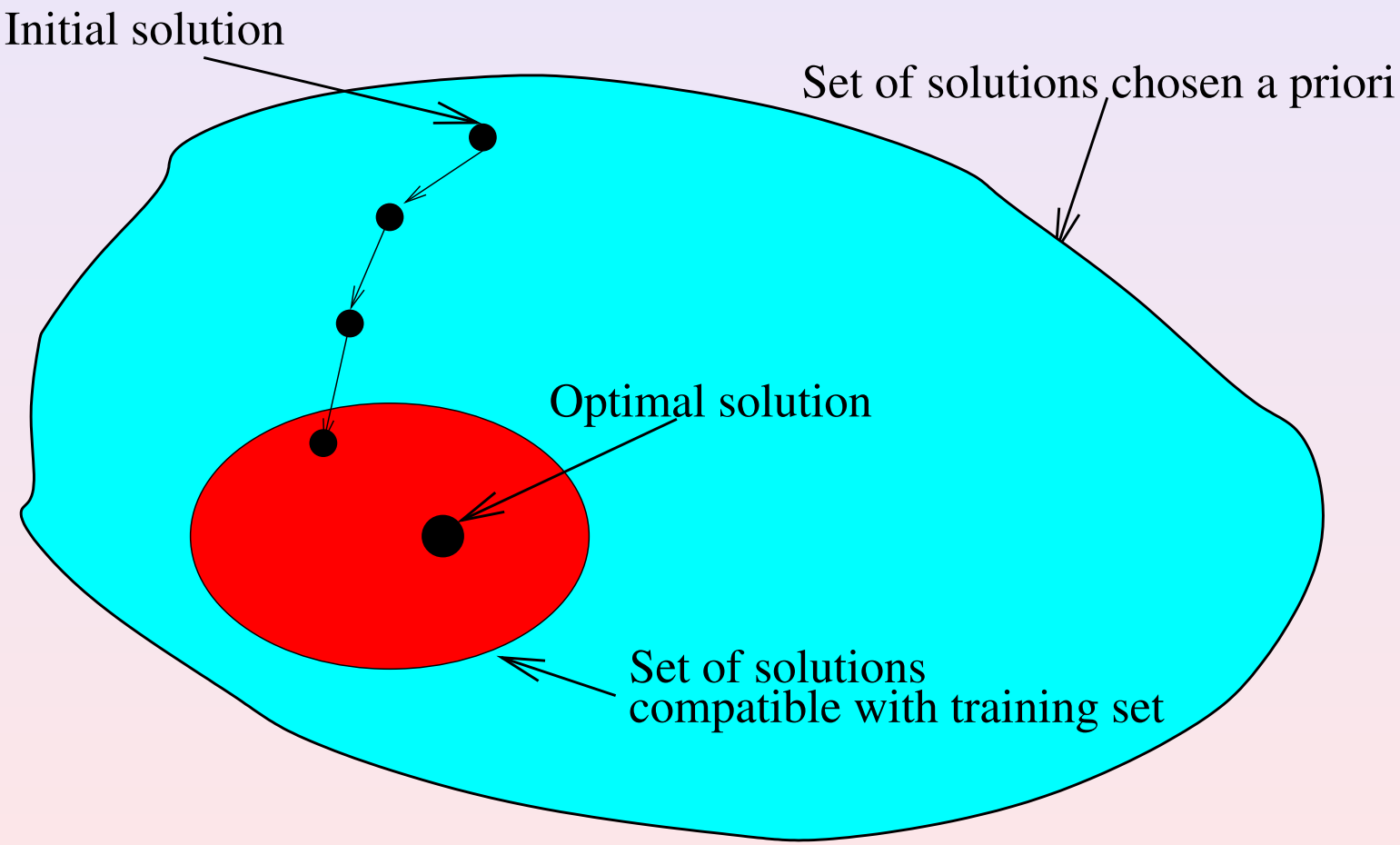
- William of Occam: Monk living in the 14th century
- Principle of Parcimony:

  One should not increase, beyond what is necessary, the number of entities required to explain anything

- When many solutions are available for a given problem, we should select the simplest one
- But what do we mean by simple?
- We will use prior knowledge of the problem to solve to define what is a simple solution

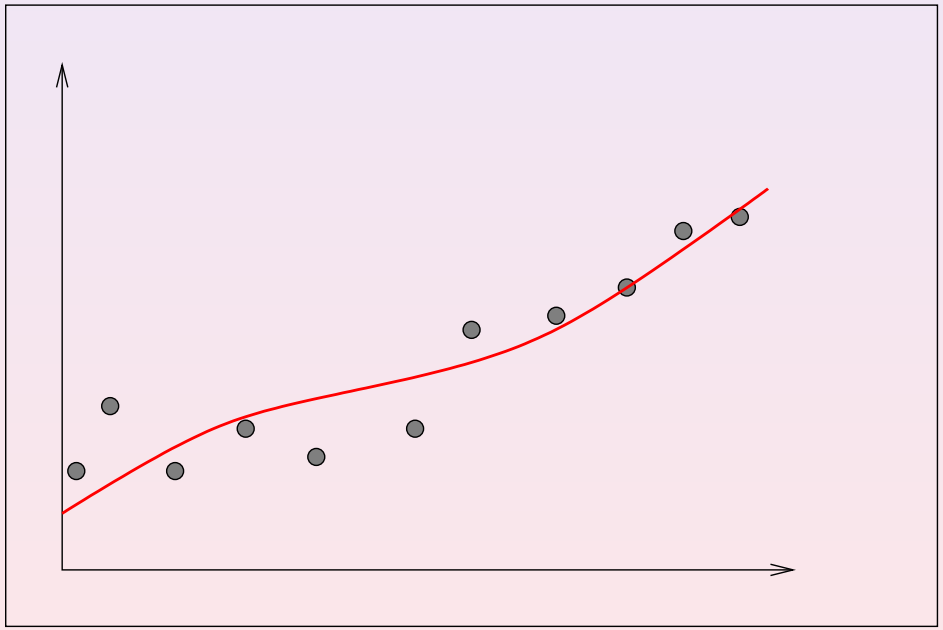
  Example of a prior: smoothness

Learning as a Search Problem



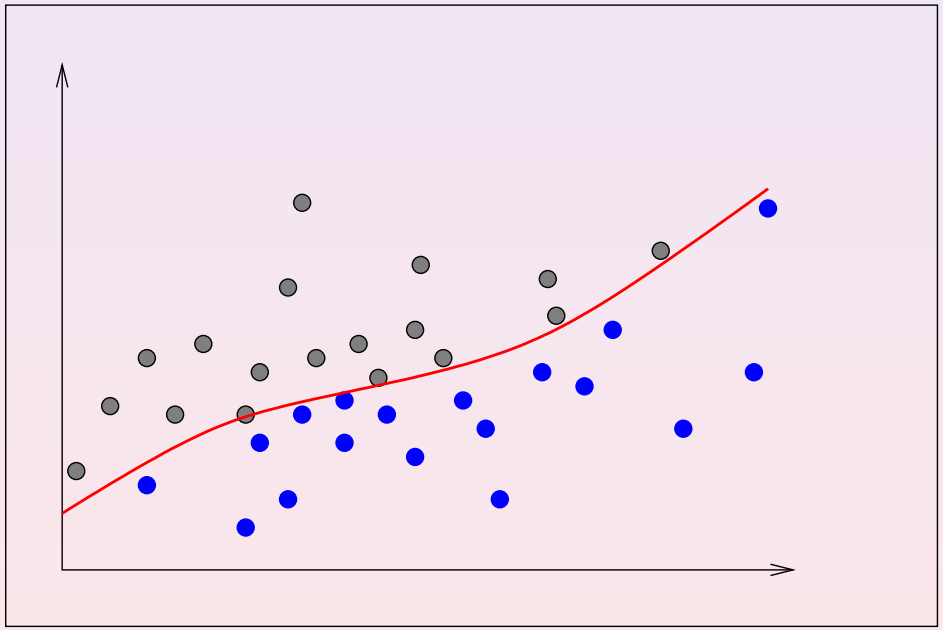
Types of Problems

- There are 3 kinds of problems:
  - regression



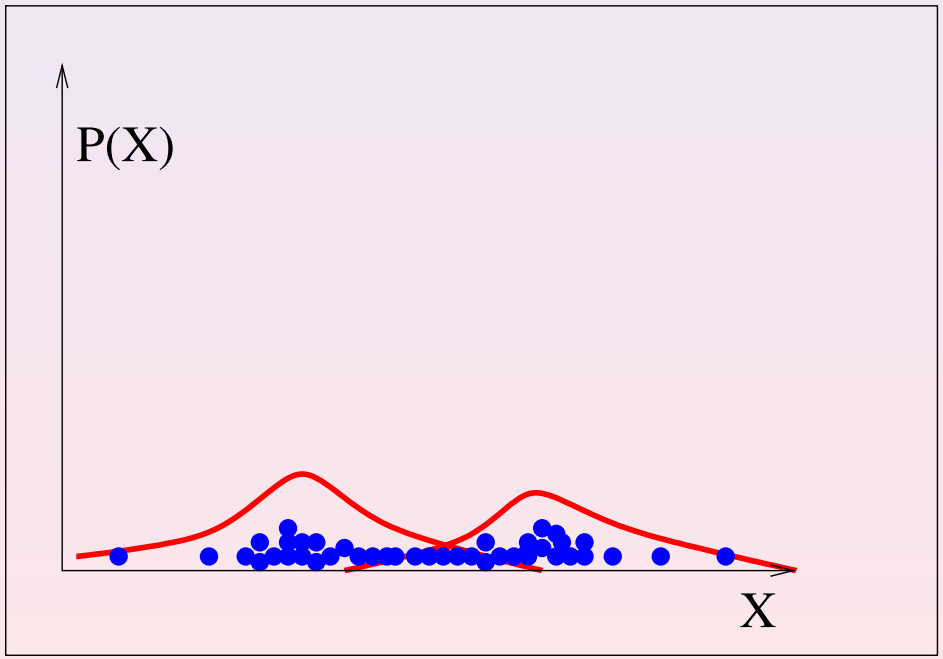
Types of Problems

- There are 3 kinds of problems:
  - regression, classification



Types of Problems

- There are 3 kinds of problems:
  - regression, classification, density estimation



Types of Learning

Supervised learning:

- The training data contains the desired behavior
- (desired class, outcome, etc)

Reinforcement learning:

- The training data contains partial targets
- (for instance, simply whether the machine did well or not).

Unsupervised learning:

- The training data is raw, no class or target is given
- There is often a hidden goal in that task (compression, maximum likelihood, etc)

Applications

- Vision Processing
  - Face detection/verification
  - Handwritten recognition
- Speech Processing
  - Phoneme/Word/Sentence/Person recognition
- Others
  - Indexing: google, text mining, information retrieval
  - Finance: asset prediction, portfolio and risk management
  - Telecom: traffic prediction
  - Data mining: make use of huge datasets kept by large corporations
  - Games: Backgammon, go
  - Control: robots
- ... and plenty of others of course!

Content of the Course

- Theoretical Issues
  - What are the theoretical foundations for statistical learning?
  - How can we measure the expected performance of a model?
- Modeling Issues
  - Models specialized for classification, regression, distributions, sequences, images, etc
  - For each model, we need to devise a training algorithm
- Others
  - Other practical issues, such as feature selection, parameter sharing, etc.
- Laboratories
  - About one third of the course will be through practical laboratories, using the python programming language

Journals and Conferences

- Journals:
  - Journal of Machine Learning Research
  - Neural Computation
  - IEEE Transactions on Neural Networks
  - IEEE Transactions on Pattern Analysis and Machine Intelligence
- Conferences:
  - NIPS: Neural Information Processing Systems
  - COLT: Computational Learning Theory
  - ICML: International Conference on Machine Learning

Books and Lecture Notes

- Books:
  - C. Bishop. Neural Networks for Pattern Recognition, 1995.
  - V. Vapnik. The Nature of Statistical Learning Theory, 1995.
  - T. Hastie, R. Tibshirani, J. Friedman. The elements of Statistical Learning, 2001.
  - B. Schölkopf, A. J. Smola. Learning with Kernels, 2002.
- Other lecture notes: (some are in french...)
  - Bengio, Y.: http://www.iro.umontreal.ca/~pift6266/A03/
  - Kegl, B.: http://www.iro.umontreal.ca/~kegl/ift6266/
  - Jordan, M.: http://www.cs.berkeley.edu/~jordan/courses/281A-fall04/
  - LeCun, Y.: http://www.cs.nyu.edu/~yann/2005f-G22-2565-001/

Electronic Resources

- Search engines:
  - NIPS online: http://nips.djvuzone.org
  - Citeseer: http://citeseer.ist.psu.edu/
  - Google scholar: http://scholar.google.com/
- Machine learning libraries:
  - Torch: http://www.Torch.ch
  - Lush: http://lush.sf.net