An Introduction to Statistical Machine Learning - Overview -

Samy Bengio

bengio@idiap.ch

Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP)

CP 592, rue du Simplon 4

1920 Martigny, Switzerland

http://www.idiap.ch/~bengio

Outline

1. Introduction to Machine Learning
2. Introduction to the Statistical Learning Theory
3. Classical Models
4. Artificial Neural Networks and Gradient Descent
5. Gaussian Mixture Models and Expectation-Maximization
6. How to Program in Torch
7. Hidden Markov Models
8. Support Vector Machines
9. Ensemble Models

Outline

10. Advanced Topics
    (a) Decision Trees
    (b) Feature Selection
    (c) Parameter Sharing
    (d) Identity Verification
    (e) Kalman Filters
    (f) Particle Filters

Documentation

- Machine learning library: www.Torch.ch
- Journals:
  - Journal of Machine Learning Research
  - Neural Computation
  - IEEE Transactions on Neural Networks
- Conferences:
  - NIPS: Neural Information Processing Systems
  - COLT: Computational Learning Theory
  - ICML: International Conference on Machine Learning
  - ICANN & ESANN: 2 European conferences
- Books:
  - Bishop, C. Neural Networks for Pattern Recognition, 1995.
  - Vapnik, V. The Nature of Statistical Learning Theory, 1995.

Documentation

- Search engines:
  - NIPS online: http://nips.djvuzone.org
  - NEC: http://citeseer.nj.nec.com/cs
- Other lecture notes: (some are in french...)
  - Bengio, Y.: http://www.iro.umontreal.ca/~bengioy/ift6266/
  - Kegl, B.: http://www.iro.umontreal.ca/~kegl/ift6266/
  - Jordan, M.: http://www.cs.berkeley.edu/~jordan/courses/294-fall98/

Introduction to Machine Learning

1. What is Machine Learning?
2. Why is it difficult?
3. Basic Principles
   (a) Occam's Razor
   (b) Learning as a Search Problem
4. Types of Problems
   (a) Regression
   (b) Classification
   (c) Density Estimation
5. Applications
6. Documentation

Statistical Learning Theory

1. The Data
2. The Function Space
3. The Loss Function
4. The Risk and the Empirical Risk
5. The Training Error
6. The Capacity
7. The Bias-Variance Dilemma
8. Regularization
9. Estimation of the Risk
10. Model Selection
11. Methodology

Classical Models

1. Parametric or Not?
2. Histograms
3. Problem: Curse of Dimensionality
4. K Nearest Neighbors
5. Parzen Windows
6. Maximum Likelihood Approach
7. Bayes Decision and Bayes Classifiers
8. K-Means

Artificial Neural Networks and Gradient Descent

1. Artificial Neural Networks
2. Multi Layer Perceptrons
3. Gradient Descent
4. ANN for Classification
5. Tricks of the Trade
6. Other ANN Models

Gaussian Mixture Models and EM

1. Reminder: Basics on Probabilities
2. What is a GMM
3. Basics of EM
4. Convergence of EM
5. EM for GMMs
6. Initialization

Hidden Markov Models

1. Markov Models
2. Hidden Markov Models
3. HMMs as Generative Processes
4. Markovian Assumptions for HMMs
5. The Likelihood given an HMM
6. EM Training for HMM
7. The Most Likely Path in an HMM
8. HMMs for Speech Recognition

Support Vector Machines

1. The aim of SVMs
2. Linear SVMs and soft margin
3. Solving the SVMs problem using a Lagrangian method
4. Kernel trick
5. Support Vector Regression

Ensemble Models

1. Basics of Ensembles
2. Bagging
3. AdaBoost

Decision Trees

1. General Model
2. Partition Function
3. Training
4. Regularization
5. Regression: CART

Feature Selection

1. Why Should We Do Select Features?
2. Broad Classes of Feature Selection
3. Wrapper Methods
4. L'autre methode!

Advanced Topic: Parameter Sharing

1. Time Delay Neural Networks
2. LeNet for Images
3. Parameter Sharing for GMMs and HMMs

Advanced Topic: Identity Verification

1. Speaker Verification
   (a) Text Dependent, Text Independent, Customized Password
   (b) Maximum Likelihood and MAP Adaptation
2. Face Verification
   (a) Localization
   (b) Verification
3. Fusion and Confidence Estimation
4. Joint Models