Big Data and Machine Learning

Krzysztof Choromanski
Google Brain Robotics
New York, NY, USA

April 30, 2017

1 Course Description

The goal of this series of lectures is to discuss some recent advances in the following fields of machine learning: kernel methods, neural networks (various generative adversarial net architectures) and reinforcement learning (with applications in robotics). Some of the discussed results will be only few-month old. In the kernel methods part the emphasis will be put on structured random feature maps methods that enable users to significantly speed up various algorithms based on kernels, provide compressed models and surprisingly, very often (as it was showed in some of the most recent papers in the field) guarantee more accurate models. Quasi Monte Carlo methods in the context of approximating RBF kernels via orthogonal transforms (instances of the structured technique) will be also discussed. In the part about neural networks we will discuss generative adversarial networks. We will consider different models, also these based on maximum mean discrepancy optimization that seem to lead to more stable architectures than the original one proposed by Ian Goodfellow and thus significantly simplify training. In the part about reinforcement learning the introduction to various reinforcement learning techniques with applications in robotics will be given. Many recent methods (some of them used in the AlphaGo algorithm) will be also discussed. In particular, we will discuss such techniques as: $TD(0)$, $TD(\lambda)$, LSTDQ, LSPI, DQN.

Remark: There is some intersection between the range of topics discussed in Data Mining classes given by the same lecturer and this one. However all the topics in this class will be discussed in greater depth so students who are/were taking Data mining classes are very welcome to attend also this one.

2 Prerequisites

The course does not require its attendees to have deep knowledge in data mining or machine learning, but some familiarity with basic constructions is useful (for instance, the lecturer assumes that the students understand some basic neural network concepts; if no then these will be explained during the lecture). All related mathematical techniques will be explained during lectures. It is also assumed that participants have some basic background in linear algebra, calculus and probability theory.

3 Grading policy

The goal of the class is to prepare students to publish papers in the field in such conferences as: ICML, NIPS and AISTATS. There will be no midterms and no final exam. Instead, student will work on projects.

4 Content

- Kernel methods with structured random feature maps (5 lectures)
  - **Lecture 1:** Introduction to structured random feature maps, concentration results for circulant matrices.
  - **Lecture 2:** Discrete structures based on Hadamard matrices and Gaussian orthogonal transforms in JLT and angular kernel approximation setting - part I.

- **Lecture 3:** Discrete structures based on Hadamard matrices and Gaussian orthogonal transforms in JLT and angular kernel approximation setting - part II.
- **Lecture 4:** Gaussian orthogonal matrices for radial basis function kernels - part I.
- **Lecture 5:** Gaussian orthogonal matrices for radial basis function kernels - part II.

- Generative Adversarial Networks (3 lectures)
  - **Lecture 6:** Introduction to generative adversarial networks - discriminators and generators, various variants of the base architecture.
  - **Lecture 7:** GANs based on maximum mean discrepancy otpimization.
  - **Lecture 8:** Compressed sensing for GANs.

- Reinforcement Learning (4 lectures)
  - **Lecture 9:** Introduction to reinforcement learning - MDP models, Bellman equations, model-based approaches.
  - **Lecture 10:** Model-free techniques: Monte-Carlo, TD-methods.
  - **Lecture 11:** Algorithms: DQN, LSTDQ, LSPI - part I.
  - **Lecture 12:** Algorithms: DQN, LSTDQ, LSPI - part II.

5 Literature:

The basis of the literature are recent papers in the field given during lectures, but some additional books that may serve as supporting material are given below.

- James, G., Witten, D. Hastie, T. and Tibshirani, R. An Introduction to Statistical Learning Springer, 2014.
- Torgo, L. Data Mining with R. CRC Press, 2011.
- Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd Edition. Springer, 2009.
- Adler, J. R in a Nutshell: A Desktop Quick Reference. OReilly Media, 2010.
- Bishop, C. Pattern Recognition and Machine Learning. Springer-Verlag, 2006.
- Witten, I. H., Frank, E. and Hall, M. A. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufman, 2011.
- Wu, X. and Kumar, V., eds. Top Ten Algorithms in Data Mining. CRC Press, 2009.

6 Office hours

Office hours will be hosted once a week in 318 Mudd, exact date will be discussed with students. Students having any questions regarding the course may send emails to: kmc2178@columbia.edu.