Seminar: Efficient Inference and Large-Scale Machine Learning

Kickoff meeting

Stephan Günnemann
Aleksandar Bojchevski
Oleksandr Shchur
Amir Moin

Technische Universität München
Department of Informatics
Data Mining and Analytics
kdd.in.tum.de

Match 1, 2017

Organization

Timeline

March

- **Until 08.03** - send us your preferred topics via email.
- **10.03** - get assigned a topic and a supervisor
- **After 10.03** - work on your topic, meet with your supervisor

April - July

- **1 week before the talk** - submission of *extended abstract* and *slides*
- **Day of the talk** - submission of *preliminary paper* for review
- **1 week after the talk** - receiving *reviews* from your peers
- **2 weeks after the talk** - submission of the *final paper*

Deliverables

Extended abstract

- 1 page, documentclass article

Paper

- 5 - 8 pages
- Latex template on the course webpage

Presentation

- 30 minutes talk
- 15 minutes discussion

Reviews

- Everyone has to review 2 papers by other students

Goal

Your paper and presentation should

- Introduce the problem setting.
- Provide a summary of the topic.
- Describe main ideas and important results.
- Mention applications and connections to other methods.

Grading

The grade is determined based on

- Extended abstract
- Report
- Presentation (slides and speech)
- Reviews written by **you**
- Involvement in the class
- Interactions with the supervisor
- Extra bonuses for own contributions (e.g. visualizations, demos, experiments)
- Penalties for missed deadlines

Background

Background: PGMs and Bayesian inference

Probabilistic graphical models and Bayesian Inference

- If you have no experience with PGMs or Bayesian Inference, it might be a good idea to get familiar with them.
- References are provided on the course webpage.

Topics

Tensor Factorization

- Generalization of matrix factorization.
- Easy to model e.g. temporal effects.
- A lot of problems (e.g. skip-gram) can be formulated as tensor factorization.

Variational Inference

- Bayesian approach is very attractive from the modeling perspective, but is often intractable.
- Cast inference problem as optimization.
- A lot of advanced ML models rely on VI.
- Recent work:
  - Black Box VI
  - Stochastic VI
  - Variational autoencoders
  - Hierarchical variational models

Message Passing and Expectation Propagation

- Inference in probabilistic graphical models.
- Minimize KL-divergence using moment matching.
- Interesting properties different from standard VI.

Particle Filters

- Approximate inference for sequential data.
- Represent belief by random samples.
- Estimation of non-Gaussian, nonlinear processes.

Sampling

- Approximate posterior inference using Monte Carlo methods.
- No constraints on the model structure.
- Recovers the true posterior in the asymptotic regime.

Natural Gradients

- Superior convergence to standard gradient optimization.
- Utilizes the structure of the optimization space.
- May result in easier to compute updates.

Bayesian Optimization

- Gradient-free optimization.
- Principled approach for hyperparameter selection.
- Related to Bayesian nonparametrics (e.g. Gaussian processes).

Probabilistic Numerics

- Interpreting numerical methods as learning algorithms.
- Efficient approximations to cost-intensive numerical procedures with uncertainty guarantees.
- Applicable to many domains (e.g. linear algebra, PDEs, uncertainty quantification).

Large-Scale Learning Systems

- Efficient implementation of all of the above.
- Scale from single CPU/GPU to massive clusters.
- Different computational paradigms (static vs. dynamic computational graph).
- Automatic differentiation.

Recap

Most important points

- Send us your preferred topics until 08.03.
- Let us know if you want to deregister until 15.03.
- Do not work on the topic completely on your own. Reach out to your supervisors.
- Brush up your background knowledge if necessary.

Questions?