Hot Topics in Machine Learning

Summer Term 2016

Prof. Marius Kloft, Florian Wenzel

April 20, 2016

Organization

Organization

The seminar is organized by Prof. Marius Kloft and Florian Wenzel (PhD student). For questions regarding the seminar please contact me:

Contact

Florian Wenzel
wenzelfl@hu-berlin.de
www.florian-wenzel.de

Organization

Course Website

http://www2.informatik.hu-berlin.de/~wenzelfl/teaching/2016_hot_topics.html

Doodle: Pick a slot please!

http://doodle.com/poll/67ffvtdnmec4qcsb

Organization

- each participant should choose (at least) one topic which she/he wants to present
- topics can be everything regarding ML (as long as it's hot)
    - interesting paper
    - interesting ML method or algorithm
    - Bachelor's or Master's thesis (work in progress is totally fine)
    - own ML project
    - choose a topic from our list of potential topics

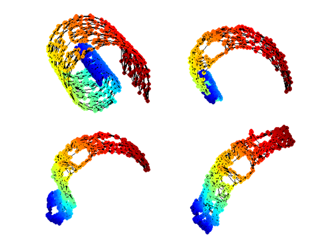
Organization

- doodle for open slots
- presentation should be around 45min + Q&A
- **2 weeks before presentation meet with Marius and discuss / rehearse presentation (~10min meeting)**
- we will meet each week (exceptions will be announced: email list)
- credit points for successful presentation and active participation

Possible Topics: Dimensionality Reduction

ISO-MAP

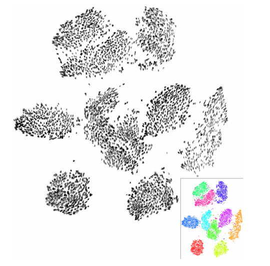
- nonlinear dimensionality reduction method
- estimate of the intrinsic geometry of a data manifold based on a rough estimate of each data point's neighbors



Sources:
`http://isomap.stanford.edu/`

t-SNE

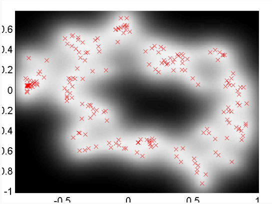
- nonlinear dimensionality reduction method
- t-SNE constructs a probability distribution over pairs of high-dimensional objects in such a way that similar objects have a high probability
- very popular and used in a wide range of applications



Sources:
http://jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

GP-LVM

- really cool nonlinear dimensionality reduction method based on Gaussian Processes
- embeds data points in a latent variable space (equipped with a prob meassure)
- gives simultaneously probabilities for data points on the learned manifold for belonging to the true (unknown) latent space



Sources:

`https://www.youtube.com/watch?v=l98Lw9KHzfc`

Paper: *gaussian process latent variable models for visualisation of high dimensional data*

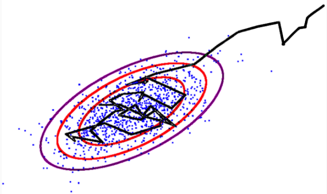
Other Dim Reduction Related Topics

- NMF (Nonnegative Matrix Factorization)
- LLE (Locally Linear Embedding)

Possible Topics: Inference

Markov Chain Monte Carlo (MCMC)

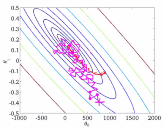
- aim: sample from a (intractable) posterior
- construct Markov chain that converges to the target distribution (as equilibrium distribution)
- for the seminar you can focus on the popular Metropolis-Hastings algorithm
- other (more advanced) MCMC algorithms: Hamiltonian Monte Carlo (HMC), SGD-based MC (next slide)



Sources:
Paper: *An Introduction to MCMC for Machine Learning*

Scalable Bayesian Inference

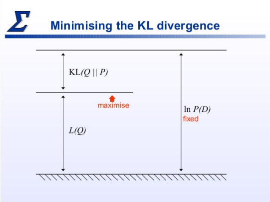
- most MCMC algorithms need swap through the whole dataset per sample
- SGD-based Sampling uses only a little fraction (so called mini batch) of the dataset for each sample
- based on Stochastic Gradient Descent (SGD)
- for seminar suitable: **SGLD (Langevin Dynamics)** or SGFS (improved version of SGLD)



Sources:
`https://www.youtube.com/watch?v=qBf5EBdEw7Q`
Paper: *Bayesian Learning via Stochastic Gradient Langevin Dynamics*
Paper: *Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring*

Variational Inference

- approximate the posterior with another (easier) distribution
- aim: Minimize the Kullback-Leibler divergence
- this is equivalent to maximizing the ELBO (evidence lower bound)
- different Assumptions lead to different algorithms, for the seminar the mean field VI algorithm is suitable
- scalable version: Stochastic Variational Inference



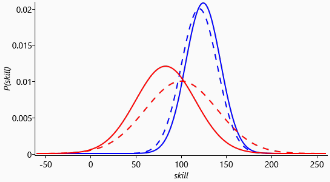
Sources:

Books: Bishop, Murphy

Paper: *Blei et al: Variational Inference: A Review for Statisticians*

Expectation Propagation

- similar idea to Variational Inference, but now minimize the reverse KL divergence
- but leads to completely different algorithm
- find approximative distribution by moment matching



Sources:

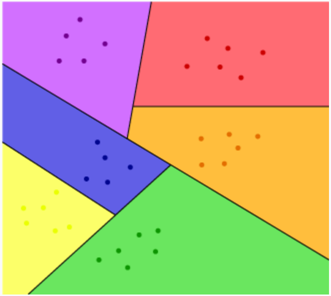
Books: Bishop, Murphy

Paper: *Minka: Expectation Propagation for Approximate Bayesian Inference*

Possible Topics: Multi Stuff

Multi Class Learning

- present different generalizations of binary class to multi class models
- compare different strategies (one-vs-rest, one-vs-one)
- focus on Multi Class SVM (present different formulations)
- extreme classification (thousands of classes)



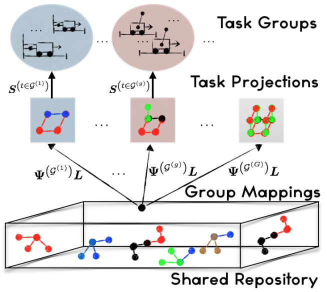
Sources:

Papers by Marius Kloft

Book: Bishop

Multi Task Learning

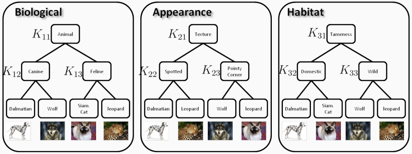
- transfer knowledge from mastering one task to the other
- idea: solve related problems at the same time, using a shared representation
- present an MTL framework (e.g. Multi Task SVM)



Sources:
Papers by Marius Kloft
Paper: Caruana: Multitask Learning

Multiple Kernel Learning

- we have a (large) set of predefined kernels and want to combine them to one
- aim: find the best weights of linear combination
- present an MKL framework (e.g. Multi Kernel SVM)
- ℓ_p-norm kernel learning (Kloft)



Sources:

PhD thesis and papers by Marius Kloft

Paper: Caruana: Multitask Learning

Possible Topics: Other Cool Possibilities

Other Topics

- CRF (Conditional Random Fields)
- Gradient Boosting
- RNNs (Recurrent Neural Networks)
- NLP Topics: Topic Models, Word Embeddings, Sentiment Analysis
- Bandits
- Online Learning Theory

Possible Topics: Your Own Ideas

Your Own Ideas

- please feel free to come up with your own topics
- explicitly welcome
- you can meet or contact me via mail if you have questions