Syllabus for the course
“Neurobayesian models”

Author	Dmitry Vetrov, Research Professor at Faculty of Computer Science NRU HSE
ECTS	3
Contact hours	34
Self-study hours	80
Study year	1
Format	Full-time

Lector: Dmitry Vetrov, Research Professor at Faculty of Computer Science NRU HSE

Tutors: Alexander Grishin, Kirill Struminsky, Dmitry Molchanov, Kirill Neklyudov, Artem Sobolev, Arsenii Ashukha, Oleg Ivanov, Ekaterina Lobacheva.

Abstract

This course is devoted to Bayesian reasoning in application to deep learning models. Attendees would learn how to use probabilistic modeling to construct neural generative and discriminative models, how to use the paradigm of generative adversarial networks to perform approximate Bayesian inference and how to model the uncertainty about the weights of neural networks. Selected open problems in the field of deep learning would also be discussed. The practical assignments will cover implementation of several modern Bayesian deep learning models.

Pre-requisites

The course is based on knowledge and understanding of Bayesian Methods and Deep Learning. It also requires programming experience in Python.

Learning Objectives

The learning objective of the course is to give students basic and advanced tools for inference and learning in complex probabilistic models involving deep neural networks, such as probabilistic deep generative models and Bayesian neural networks.

Learning Outcomes

After completing the study of the discipline, the student should have:

- Knowledge about different approximate inference and learning techniques for probabilistic models
- Hands-on experience with modern probabilistic modifications of deep learning models
- Knowledge about the necessary building blocks that allow to construct new probabilistic models, suitable for the desired problems.

Course Plan:

1. ***Stochastic Variational Inference (SVI)***
   SVI as a scalable alternative to the variational inference for tasks with large data. Application of SVI to latent Dirichlet allocation model.

2. ***Variational autoencoders (VAE) and normalizing flows***
   Probabilistic PCA, VAE as a non-linear generalization of probabilistic PCA. Reparametrization trick for doubly-stochastic variational inference. Extending variational approximations with normalizing flows. Examples of normalizing flows.
3. ***Implicit Variational Inference using Adversarial Training***
   Adversarial Variational Bayes for training VAE with implicit inference distribution. f-GANs as a generalization of vanilla GANs for optimizing arbitrary f-divergence.
4. ***Bayesian neural networks***
   Variational inference of the posterior distribution over the weights of discriminative neural networks. Local reparameterization trick for gradient variance reduction.
5. ***Bayesian compression of neural networks***
   Variational Dropout sparsifies deep neural networks: different parametrization yields totally different model. Soft Weight Sharing: how to save memory, using weights quantization of neural network.
6. ***Deep MCMC***
   How neural networks help MCMC methods to sample from analytical distribution, and how MCMC methods help neural networks to sample from empirical distribution.
7. ***Discrete Latent Variables and Variance Reduction***
   The idea of Stochastic Computation Graphs, discrete and continuous stochastic nodes, and gradient estimation: Gumbel-Softmax and REINFORCE with control variates.
8. ***Semi-implicit variational inference***
   Semi-implicit probabilistic models as a relaxation of implicit models, that can allow for more efficient approximate inference. Implicit optimal priors. Implicit aggregated posterior as an optimal prior distribution for VAEs and several ways to approximate it and use it in inference.

Grading System

The assessment consist of 3 practical assignments and a final oral exam. Practical assignments consist in programming some models/methods from the course in Python and analysing their behavior: VAE, Normalizing flows, Sparse Variational Dropout. At the final exam students have to demonstrate knowledge of the material covered during the entire course.

Final course grade is obtained from the following formula:

$$O\_final = 0{,}7 * O\_cumulative + 0{,}3 * O\_exam,$$

where *O_cumulative* is an average grade for the practical assignments.

All grades are in ten-point grading scale. If *O_cumulative* or *O_final* has a fractional part greater or equal than 0.5 then it is rounded up. Each practical assignment has a deadline, a penalty is charged in the amount of 0.3 points for each day of delay, but in total not more than 6 points. Students have to complete all assignments by themselves, plagiarism is strictly prohibited.

Reading List

Books:

Murphy K.P. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
Bishop C.M. Pattern Recognition and Machine Learning. Springer, 2006.
Mackay D.J.C. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.
Ian Goodfellow, Yoshua Bengio & Aaron Courville. Deep Learning. MIT Press, 2016.

Papers:

Stochastic Variational Inference (SVI)

Matthew D. Hoffman et al. Stochastic Variational Inference. JMLR 2013. http://jmlr.org/papers/volume14/hoffman13a/hoffman13a.pdf

Variational autoencoders (VAE) and normalizing flows

Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. ICLR 2014. https://arxiv.org/abs/1312.6114

Danilo Jimenez Rezende and Shakir Mohamed. Variational Inference with Normalizing Flows. ICML 2015. https://arxiv.org/abs/1505.05770

Implicit Variational Inference using Adversarial Training

Lars Mescheder et al. Adversarial Variational Bayes. ICML, 2017. https://arxiv.org/abs/1701.04722

Sebastian Nowozin et al. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. NIPS, 2017. https://arxiv.org/abs/1606.00709

Bayesian neural networks

Blundell, Charles, et al. Weight uncertainty in neural networks. ICML 2015. https://arxiv.org/abs/1505.05424

Wang, Sida, and Christopher Manning. Fast dropout training. ICML 2013. http://proceedings.mlr.press/v28/wang13a.pdf

Kingma, Durk P. et al. Variational dropout and the local reparameterization trick. NIPS 2015. https://arxiv.org/abs/1506.02557

Louizos, Christos, and Max Welling. Multiplicative normalizing flows for variational bayesian neural networks. ICML 2017. https://arxiv.org/abs/1703.01961

Bayesian compression of neural networks

Dmitry Molchanov et al. Variational Dropout Sparsifies Deep Neural Networks. ICML 2017. https://arxiv.org/abs/1701.05369

Karen Ullrich et al. Soft Weight-Sharing for Neural Network Compression. ICLR 2017. https://arxiv.org/abs/1702.04008

Deep MCMC

Jiaming Song et al. A-NICE-MC: Adversarial Training for MCMC. NIPS 2017. https://arxiv.org/abs/1706.07561

Daniel Levy et al. Generalizing Hamiltonian Monte Carlo with Neural Networks. ICLR 2018. https://arxiv.org/abs/1711.09268

Discrete Latent Variables and Variance Reduction

Eric Jang et al. Categorical Reparameterization with Gumbel-Softmax. ICLR 2017. https://arxiv.org/abs/1611.01144

Chris J. Maddison et al. The Concrete Distribution: A Continuous Relaxation of Discrete Random Variables. ICLR 2017. https://arxiv.org/abs/1611.00712

George Tucker et al. REBAR: Low-variance, unbiased gradient estimates for discrete latent variable models. NIPS 2017. https://arxiv.org/abs/1703.07370

Will Grathwohl et al. Backpropagation through the Void: Optimizing control variates for black-box gradient estimation. ICLR 2018. https://arxiv.org/abs/1711.00123

Semi-implicit variational inference

Yin, Mingzhang, and Mingyuan Zhou. Semi-Implicit Variational Inference. ICML 2018. https://arxiv.org/abs/1805.11183

Tomczak, Jakub M., and Max Welling. VAE with a VampPrior. AISTATS 2018. https://arxiv.org/abs/1705.07120

Hiroshi Takahashi et al. Variational Autoencoder with Implicit Optimal Priors. arXiv preprint 2018. https://arxiv.org/abs/1809.05284

Dmitry Molchanov et al. Doubly Semi-Implicit Variational Inference." AISTATS 2019. https://arxiv.org/abs/1810.02789