Bayesian Neural Networks

ROY team: Ilya Zharikov,
Roman Isachenko,
Artem Bochkarev

Skolkovo Institute of Science and Technology
Bayesian Methods course

May 25, 2017

Project goal

Aim

Estimate posterior distribution of the model parameters from data

Problem

Monte Carlo sampling is very slow for high-dimensional data

Probabilistic Programming

- Uncertainty in predictions
- Uncertainty in representations
- Regularizations with priors
- Transfer learning
- Hiearchical Neural Networks

Related work

1. Salvatier J, Wiecki T. V., Fonnesbeck C. Probabilistic programming in Python using PyMC3. // *PeerJ Computer Science*. 2016.
2. Blundell C. et al. Weight Uncertainty in Neural Network // *Proceedings of The 32nd International Conference on Machine Learning*. 2015.
3. Kucukelbir A. et al. Automatic Differentiation Variational Inference // *arXiv preprint arXiv:1603.00788*. – 2017.

Problem Statement

Bayes Theorem

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}$$

Frequentist approach

$$\theta^* = \arg\max_{\theta} p(\theta|X) = \arg\max_{\theta} p(X|\theta) + p(\theta)$$

Monte Carlo approach

- Metropolis Hasting
- Gibbs sampling
- No-U-Turn Sampling (NUTS)

Variational Inference

$$\ln p(X) = \mathrm{KL}(q||p) + \mathrm{ELBO}(q)$$

$$\mathrm{KL}(q||p) = \int q(\theta) \ln \frac{q(\theta)}{p(\theta|X)} d\theta; \quad \mathrm{ELBO}(q) = \int q(\theta) \ln \frac{p(X,\theta)}{q(\theta)} d\theta$$

Problem

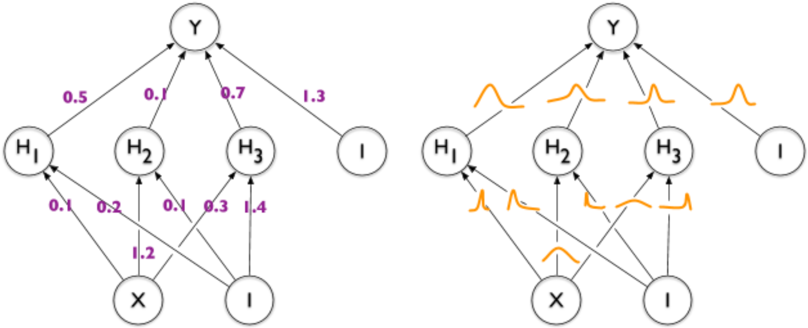
minimization of $\mathrm{KL}(q||p) \Leftrightarrow$ maximization of $\mathrm{ELBO}(q)$

ADVI

- Transformation of constrained variables
- $q(\theta)$ comes from parametric family (usually $\mathcal{N}(\mu, \mathrm{diag}(\sigma^2))$)
- Stochastic optimization
- Integral differentiation $\Rightarrow$ reparametrization trick

Neural Networks

- **Neural networks** predict values of parameters by fitting complex model on the huge dataset
- **Bayesian Neural Networks** predict the parameters distributions



http://bit.ly/2rMQuDq