Variational Inference and Deep Generative Models

Wilker Aziz
ILLC
University of Amsterdam
w.aziz@uva.nl

Philip Schulz
Amazon Research
phschulz@amazon.com

1 Tutorial Contents

Neural networks are taking NLP by storm. Yet they are mostly applied to fully supervised tasks. Many real-world NLP problems require unsupervised or semi-supervised models, however, because annotated data is hard to obtain. This is where generative models shine. Through the use of latent variables they can be applied in missing data settings. Furthermore they can complete missing entries in partially annotated data sets.

This tutorial is about how to use neural networks inside generative models, thus giving us Deep Generative Models (DGMs). The training method of choice for these models is variational inference (VI). We start out by introducing VI on a basic level. From there we turn to DGMs. We justify them theoretically and give concrete advise on how to implement them. For continuous latent variables, we review the variational autoencoder and use Gaussian reparametrisation to show how to sample latent values from it. We then turn to discrete latent variables for which no reparametrisation exists. Instead, we explain how to use the score-function or REINFORCE gradient estimator in those cases. We finish by explaining how to combine continuous and discrete variables in semi-supervised modelling problems.

2 Schedule

1. Introduction (20 minutes)
   - Maximum likelihood learning
   - Stochastic gradient estimates
   - Unsupervised learning
2. Basics of Variational Inference (45 minutes)
   - Review of posterior inference and intractable marginal likelihoods
   - NLP examples
   - Derivation of variational inference
   - Mean field approximation

20 minutes break

3. DGMs with Continuous Latent Variables (45 minutes)
   - Wake-sleep algorithm
   - Variational autoencoder
   - Gaussian reparametrisation
4. DGMs with Discrete Latent Variables (30 minutes)
   - Latent factor model
   - Why discrete variables cannot be reparametrised
   - Score function gradient estimator
   - Comparison of reparametrisation and score function estimators
   - Semi-supervised learning
5. Q&A

3 About the Presenters

Wilker Aziz is a research associate at the University of Amsterdam (UvA) working on natural language processing problems such as machine translation, textual entailment, and paraphrasing. His research interests include statistical learning, probabilistic models, and methods for approximate inference. Before joining UvA, Wilker worked on exact sampling and optimisation for statistical machine translation at the University of Sheffield (UK) and at the University of Wolverhampton (UK) where he obtained his PhD. Wilker's background is in Computer Engineering which he studied at the Engineering School of the University of São Paulo (Brazil).

Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics-Tutorial Abstracts, pages 8–9
Melbourne, Australia, July 15 - 20, 2018.

Philip Schulz is an applied scientist at Amazon Research. Before joining Amazon, Philip did his PhD at the University of Amsterdam. During the last months of his PhD trajectory, he visited the University of Melbourne. Philip's background is in Linguistics which he studied at the University of Tübingen and UCL in London. These days, his research interests revolve around statistical learning. He has worked on Bayesian graphical models for machine translation. More recently he has extended this line of work towards deep generative models. More broadly, Philip is interested in probabilistic modeling, approximate inference methods and statistical theory.