UPPSALA UNIVERSITET

Stochastic Backpropagation and Approximate Inference in Deep Generative Models

Carl Andersson

Uppsala University, Department of Information Technology

UPPSALA UNIVERSITET

Introduction

Why probabilistic models?

- Prediction, what comes next in a data sequence
- Data imputation, fill in an image or a data record with probable information
- Uncertainty estimation, not only a single value but a distribution

carl.andersson@it.uu.se

Introduction

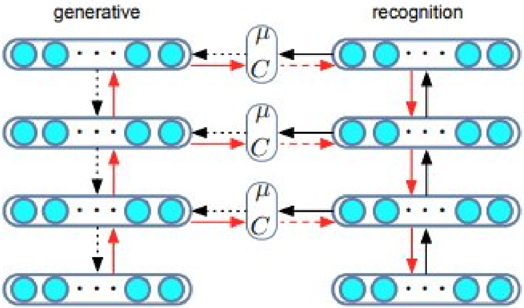
Core concepts

- Deep architecture, be able to represent complex structures in the data
- Fast sampling, not needing any tedious sampling step such as Gibbs sampling
- Tractable and scalable, applicable to high dimensional data

carl.andersson@it.uu.se

Deep architecture

Stacking layers of latent variables and let them depend through simple, single layered neural networks



UPPSALA UNIVERSITET

Previous works

Models

- Factor analysis
- Nonlinear Gaussian belief networks

Inference

- Mean-field variational EM
- Wake-sleep algorithm
- Stochastic variational+control variates (Last paper)

Variational Autoencoder, Kingma & Welling

carl.andersson@it.uu.se

Stochastic Backpropagation

A way to differentiate through a random variable
eg. Gaussian Backpropagation

$$\nabla_\theta \mathbb{E}_{\mathcal{N}(\xi|\mu,C)}[f(\xi)] = \mathbb{E}_{\mathcal{N}(\xi|\mu,C)}\left[\frac{\partial \mu}{\partial \theta}\nabla f + \frac{1}{2}\mathrm{Tr}\left(\frac{\partial C}{\partial \theta}\nabla^2 f\right)\right]$$

- Requires the Hessian ☹

Reparametrization trick

$$\nabla_\theta \mathbb{E}_{\mathcal{N}(\xi|\mu,C)}[f(\xi)] = \mathbb{E}_{\mathcal{N}(\epsilon|0,1)}\left[\nabla_\theta f(\mu + \epsilon R)\right]$$

carl.andersson@it.uu.se

Free energy

Maximize Evidence lower bound, ELBO

ELBO = data likelihood + KL(approximate posterior, model)

carl.andersson@it.uu.se

Further reading

Ladder Variational Autoencoders, Sønderby et al. 2016
Wasserstein GAN Arjovsky et al. 2017