

Modules

last modified: June 13, 2017

The modules below are in no particular order (except for the Basics, of course).¹

1 Basics

- What is a posterior **and inference**? → recap of Bayes' rule
- Example problems: Factorial HMMs, Bayesian Mixture Models (show GMs)
- **Is it worth to use sampling as an intuitive way of performing inference before diving in the realms of VI?**
- ELBO derivation I: from KL divergence
- ELBO derivation II: with Jensen's inequality
- **Connection to EM**
- Mean Field inference
- Application to example problems (show GMs)

2 Conjugate Models

- Exponential families
- Gaussian-Gaussian conjugacy
- Example: Gaussian Mixture Model
- **Beta-Binomial warmup?**
- Dirichlet-multinomial conjugacy

¹The order is actually not too arbitrary :) I noted a few exceptions below. Let's try and come up with an ideal order or some other form of dependency.

- Example: Multinomial Mixture Model (maybe even LDA? **Yes!**)
- Conjugate VI in the general case

3 Nonconjugate Models

- Laplace Approximation (**depending on audience**)
- Gradient methods
- Problem: cannot simply differentiate an MC average
- Idea: transform $\frac{d}{dq} \mathbb{E}_q[\cdot]$ into $\mathbb{E}[\frac{d}{dq} \cdot]$
- Score function gradient \rightarrow Black Box VI
- Reparametrisation gradient

4 Nonparametric Models

I would push this module further down and depending on audience drop it altogether. I suppose people who do not know VI, do not know nonparametric models either. Thus we would need to invest quite some time in this.

- Intro to stick-breaking processes
- VI for HDP/PYP
- Intro to GPs
- VI for GPs

5 Bayesian Neural Networks

- Putting priors on weights
- The old stuff by Neal, MacKay and Hinton
- The new stuff by DeepMind
- Bayesian Interpretation of Dropout

6 Deep Generative Models

- Review of generative models
- Exact case: EM with features
- Variational Autoencoders
- Example models: ???
- Code snippet ??? We can aim to have notebooks for some (or all?) of the modules.
- Extra: The Deep Generative CRF (the Ryan Adams paper from NIPS)

7 Reparametrisation Gradients

I think the whole module should depend on audience and we can cover the location-scale case in the modules about Nonconjugate models and/or DGMs.

- Recap: Gaussian reparametrisation
- Extension to general location-scale families
- ADVI (depending on the audience only go until here; the next two are way more complicated)
- Generalised Reparametrisation Gradient
- Rejection Sampling VI

8 Beyond Mean Field [Advanced]

- Structured VI (example: Bayesian or Factorial HMMs)
- Auxiliary variables
- Hierarchical Variational models

9 Collapsed VB

Another module that depends on audience: people with Bayesian aspirations vs people who want to play with DGMs.

- Taylor expansions
- Example: LDA

- Connection between collapsed VB and unconstrained variational approximation
- CVB0

10 Beyond KL [Advanced+]

- α -divergence (make connection to EP)
- Stein VI
- Implicit models
- Hoelder bound

11 Not sure where to fit

- Stochastic optimisation: at least at a high level
- GAN: if Eric Xing's connection between VAEs and GANs turn out interesting
- I note that NLP2 students (and colleagues of mine) struggle to understand what it means to impose a prior. We can try to clear that out (perhaps in module Conjugate Models).
- People are usually ready to quote “regularisation is an approximate Bayesian prior” but they do not understand the limits/implications of the word “approximate” there and in a way they perceive it as not too different from “VI is approximate posterior inference”. Perhaps this is worth discussing when we talk about Bayesian interpretations of (stochastic) regularisation techniques in the module BNNs.