Modules

last modified: June 13, 2017

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

1 Basics

- What is a posterior and what is posterior inference? → recap of Bayes' rule
- Sampling as an intuitive way of performing inference before diving in the realms of VI?
- Example problems: Factorial HMMs, Bayesian Mixture Models (show GMs)
- 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: Bayesian Linear Regression
- Beta-Binomial warmup for Dirichlet-multinomial?
- Dirichlet-multinomial conjugacy
- Example: Multinomial Mixture Model (and LDA)
- Conjugate VI in the general case (Beal, 2003)

3 Nonconjugate Models

- Laplace Approximation
- 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 (Blei et al., 2012; Ranganath et al., 2014)
- Reparametrisation gradient (Kingma and Welling, 2013; Rezende et al., 2014; Titsias and Lzaro-gredilla, 2014)

4 Nonparametric Models

I would put this module as advanced.

- Intro to stick-breaking processes (Ishwaran and James, 2001)
- VI for HDP/PYP (Wang et al., 2011)
- Intro to GPs
- VI for GPs

5 Bayesian Neural Networks

- Putting priors on weights
- The old stuff by Neal, MacKay and Hinton (Hinton and van Camp, 1993)
- The new stuff by DeepMind et al. (Graves, 2011; Blundell et al., 2015)
- Bayesian Interpretation of Dropout (Gal, 2016)

6 Deep Generative Models

- Review of generative models
- Exact case: EM with features (Berg-Kirkpatrick et al., 2010)
- First attempt: Wake-sleep (Hinton et al., 1995)

- Variational Autoencoders (Kingma and Welling, 2013; Rezende et al., 2014)
- Example models: ???
- Code snippet ???
- 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
- Exension to general location-scale families (Titsias and Lzaro-gredilla, 2014)
- ADVI (depending on the audience only go until here; the next two are way more complicated) (Kucukelbir et al., 2017)
- Generalised Reparametrisation Gradient (Ruiz et al., 2016)
- Rejection Sampling VI (Naesseth et al., 2017)

8 Beyond Mean Field [Advanced]

- Structured VI (example: Bayesian or Factorial HMMs)
- Auxiliary variables
- Hierarchical Varational 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 (Teh et al., 2007)
- CVB0 (Asuncion et al., 2009)

10 Beyond KL [Advanced]

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

11 Not sure where to fit

- Stochastic optimisation (Robbins and Monro, 1951): 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.

References

Arthur Asuncion, Max Welling, Padhraic Smyth, and Yee Whye Teh. On smoothing and inference for topic models. UAI '09, pages 27–34, 2009. URL https://arxiv.org/pdf/1205.2662.pdf.

Matthew J. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003. URL http://www.cse.buffalo.edu/faculty/mbeal/thesis/index.html.

Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero, and Dan Klein. Painless unsupervised learning with features. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 582–590, 2010. URL http://www.aclweb.org/anthology/N10-1083.

David M. Blei, Michael I. Jordan, and John W. Paisley. Variational bayesian inference with stochastic search. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 2012. URL http://icml.cc/2012/papers/687.pdf.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. ICML'15, pages 1613–1622, 2015. URL http://machinelearning.wustl.edu/mlpapers/paper_files/icml2015_blundell15.pdf.

Yarin Gal. *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge, 2016. URL http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf.

Alex Graves. Practical variational inference for neural networks. In *NIPS*, pages 2348–2356, 2011. URL http://papers.nips.cc/paper/4329-practical-variational-inference-for-neural-networks.

G. E. Hinton, P. Dayan, B. J. Frey, and R. M. Neal. The wake-sleep algorithm for unsupervised neural networks. *Science*, 268:1158–1161, 1995. URL http://www.gatsby.ucl.ac.uk/~dayan/papers/hdfn95.pdf.

Geoffrey E. Hinton and Drew van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, COLT '93, pages 5–13. ACM, 1993. ISBN 0-89791-611-5. URL http://doi.acm.org/10.1145/168304.168306.

Hemant Ishwaran and Lancelot F James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453): 161–173, 2001. doi: 10.1198/016214501750332758. URL http://dx.doi.org/10.1198/016214501750332758.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. 2013. URL http://arxiv.org/abs/1312.6114.

Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M. Blei. Automatic differentiation variational inference. *Journal of Machine Learning Research*, 18(14):1–45, 2017. URL http://jmlr.org/papers/v18/16-107.html.

Christian Naesseth, Francisco Ruiz, Scott Linderman, and David Blei. Reparameterization gradients through acceptance-rejection sampling algorithms. AISTATS, pages 489–498, 2017. URL http://proceedings.mlr.press/v54/naesseth17a/naesseth17a.pdf.

Rajesh Ranganath, Sean Gerrish, and David Blei. Black Box Variational Inference. In Samuel Kaski and Jukka Corander, editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pages 814–822, 2014. URL http://proceedings.mlr.press/v33/ranganath14.pdf.

Danilo J. Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1278–1286, 2014. URL http://jmlr.org/proceedings/papers/v32/rezende14.pdf.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, September 1951. URL http://projecteuclid.org/download/pdf_1/euclid.aoms/1177729586.

Francisco R Ruiz, Michalis Titsias RC AUEB, and David Blei. The generalized reparameterization gradient. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 460–468. 2016. URL http://papers.nips.cc/paper/6328-the-generalized-reparameterization-gradient.pdf.

Y. W. Teh, D. Newman, and M. Welling. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In *Advances in Neural Information Processing Systems*, volume 19, 2007. URL https://www.stats.ox.ac.uk/~teh/research/inference/nips2006.pdf.

Michalis Titsias and Miguel Lzaro-gredilla. Doubly stochastic variational bayes for non-conjugate inference. In Tony Jebara and Eric P. Xing, editors, *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1971–1979, 2014. URL http://jmlr.org/proceedings/papers/v32/titsias14.pdf.

Chong Wang, John Paisley, and David M. Blei. Online variational inference for the hierarchical dirichlet process. In *Proc. of the 14th Int'l. Conf. on Artificial Intelligence and Statistics (AISTATS)*, volume 15, pages 752–760, 2011. URL http://jmlr.csail.mit.edu/proceedings/papers/v15/wang11a/wang11a.pdf.