Homework 5: Bayesian Graphical Models

Instructions:

1. You may discuss this assignment with other students in the class, but you must submit your own answers to the questions below.
2. Include an honor pledge with your submission.
3. Submit on-line.
4. This homework is worth 100 points and the point totals for each question are shown in parentheses.

Assignment:

1. (30) This problem explores the use of variational approximation in Latent Dirichlet Allocation (LDA). We will use the implementation in sklearn of the variational approximation algorithm in [1].
   (a) Use the notation in the diagram in Figure 1 [2] to write the target posterior distribution of the latent variables and parameters for the general LDA method. Why do we use variational approximation rather than conjugate priors or sampling to obtain this posterior distribution?
   (b) Accident reports provide a good use-case for LDA since the narrative information in these reports is frequently overlooked in safety analysis. LDA allows us to capture elements (topics) in this narrative data and use them to better understand unsafe conditions. For this use-case, modify the LDA class for Wikipedia in the `LDA Examples - Wikipedia and Trains` jupyter notebook to perform LDA on the accident narratives. About 10 years of these narratives are in the json file, `TrainNarratives.txt`. Use this class to obtain 10 topics from the accident narratives.
   (c) Use the class you developed for Problem 1b to obtain the probabilities for each of the topics in the first 10 narratives in the `TrainNarratives.txt` data set. What is the notation in Figure1 that represents these probabilities?
   (d) Briefly explain how a safety engineer at Federal Railroad Administration could use the results you obtain in Problem 1c to improve safety for trains.

2. (10) Neural networks are a graphical model, Markov Networks, that can be analyzed with Bayesian methods and Boltzmann machines are examples. Why is training or learning computationally challenging for Boltzmann machines? Explain how restricted Boltzmann machines (RBM) are easier to train. What approaches do we use for training RBM?

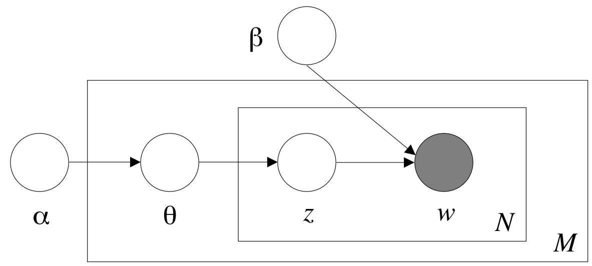

Figure 1: Graphical Model of the LDA [2]

3. (30) The following questions are based on reading and running the jupyter notebook, `pymc3-variation-inference-neural-network.ipynb`, by Thomas Wiecki, updated by Maxim Kochurov as provided in their blog post. Run the notebook and then answer these questions.

   (a) Wieki says that an advantage to using Bayesian modeling with neural network and deep learning is that "we could train the model specifically on samples it is most uncertain about." Explain how he finds these samples in this example. Explain how you would implement his suggestion (you do not have to actually implement this).

   (b) Wieki also says that another advantage to Bayesian modeling with neural network and deep learning is that "We also get uncertainty estimates of our weights which could inform us about the stability of the learned representations of the network." Discuss what the uncertainty estimates for the weights found for the example in this notebook imply.

   (c) Explain how the Gaussian priors help to regularize the weights in the neural network.

   (d) Why do we use a variational approximation instead of sampling for estimating the posterior of the weights?

   (e) Change the prior distributions for all three sets of the neural net weights to Cauchy with location (alpha) = 0 and scale (beta) = 2. Rerun the remaining cells in the notebook and comment on any changes you see from this.

4. (30) You are tracking the performance of a set of companies with the idea that you might possibly buy stock in them. You decide to automate this process using HMM and you implement your first version for one company. This company has three states that are hidden from investors: (1) in-trouble; (2) static; and (3) major growth potential. You have estimated the transition probabilities between states as follows:

$$\begin{bmatrix} .6 & .3 & .1 \\ .4 & .4 & .2 \\ .1 & .4 & .5 \end{bmatrix}$$

You have a text analysis system using Naive Bayes to process the quarterly reports and assess their sentiment into one of three categories: (1) Fine; (2) Good; and (3) Very good. Your estimates for the probabilities of these sentiments given the state of the company are shown in the following matrix (the sentiments are in the rows and the states of the company are in the columns).

$$\begin{bmatrix} .45 & .4 & .15 \\ .3 & .4 & .3 \\ .2 & .5 & .3 \end{bmatrix}$$

You have 3 quarterly reports with the assessments: Fine, Fine, Very Good and your prior for the initial state is equally likely for each value. The following questions use the HMM class in the jupyter notebook, `HMM Examples HW5 - Burglary and Investment` with your additions to it as indicated in the notebook.

(a) In order to decide whether to invest, find the most likely current state given the observed states.

(b) Using smoothing to find the most likely state at each previous time period (i.e., periods 1 and 2).

(c) Show the most likely path of performance through the hidden states up to the current time.

(d) Find the most likely hidden state and visible state for this company in the next time period.

References

[1] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, "Stochastic variational inference," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1303–1347, 2013.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.