

CPSC 540 Assignment 4 (due Friday March 27 at midnight)

1. Name(s):
2. Student ID(s):

1 Bayesian Inference

Consider a $y \in \{1, 2, 3, \dots\}$ following a Poisson distribution with rate parameter $\lambda > 0$,

$$p(y | \lambda) = \frac{\lambda^y \exp(-\lambda)}{y!}.$$

We'll assume that λ follows a Gamma distribution (the conjugate prior to the Poisson) with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$,

$$\lambda \sim \text{Gamma}(\alpha, \beta),$$

or equivalently that

$$p(\lambda | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda),$$

where Γ is the gamma function.

Compute the following quantities:

1. The posterior distribution,

$$p(\lambda | y, \alpha, \beta).$$

2. The marginal likelihood of y given the hyper-parameters α and β ,

$$p(y | \alpha, \beta) = \int p(y, \lambda | \alpha, \beta) d\lambda.$$

3. The posterior mean estimate for λ ,

$$\mathbb{E}_{\lambda | y, \alpha, \beta}[\lambda] = \int \lambda p(\lambda | y, \alpha, \beta) d\lambda.$$

4. The posterior predictive distribution for a new independent observation \tilde{y} given y ,

$$p(\tilde{y} | y, \alpha, \beta) = \int p(\tilde{y}, \lambda | y, \alpha, \beta) d\lambda.$$

where $\alpha^{++} = \tilde{y} + \alpha^+$ and $\beta^{++} = 1 + \beta^+$. Note that this is the marginal likelihood of the new data, if we treat the posterior we got from the old data as our prior (which is intuitive if you think about it).

Hint: You should be able to use the form of the gamma distribution to solve all the integrals that show up in this question. You can use $Z(\alpha, \beta) = \frac{\Gamma(\alpha)}{\beta^\alpha}$ to represent the normalizing constant of the gamma distribution, and use α^+ and β^+ as the updated parameters of the gamma distribution in the posterior.

2 Markov Chain Monte Carlo

If you run `example_MH.jl`, it loads a set of images of ‘2’ and ‘3’ digits. It then runs the Metropolis MCMC algorithm to try to generate samples from the posterior over w , in a logistic regression model with a Gaussian prior. Once the samples are generated, it makes a histogram of the samples for several of the variables.¹

1. Why would the samples coming from the Metropolis algorithm not give a good approximation to the posterior?
2. Modify the proposal used by the demo to $\hat{w} \sim \mathcal{N}(w, (1/100)I)$ instead of $\hat{w} \sim \mathcal{N}(w, I)$. Hand in your code and the update histogram plot.
3. Modify the proposal to use $\hat{w} \sim \mathcal{N}(w, (1/10000)I)$. Do you think this performs better or worse than the previous choice? (Briefly explain.)

3 Deep Structured Models

Cancelled.

4 Very-Short Answer Questions

Give a short and concise 1-sentence answer to the below questions.

1. Under what conditions can we use alpha-beta swap moves for approximate decoding?
2. Why do we use the parameterization $\phi_j(s) = \exp(w_{j,s})$ in UGMs?
3. What is the key difference between Younes’ algorithm and our usual setting of using an unbiased approximation of the gradient?
4. What is the key advantage of the graph structure in restricted Boltzmann machines?
5. What is the advantage of using a CRF, modeling $p(y | x)$, rather than treating supervised learning as special case of density estimation (modeling $p(y, x)$).
6. Why can fully-convolutional networks segment images of different sizes?
7. What is the key feature of a “sequence to sequence” RNN?
8. What are two advantages of the Bayesian approach to learning?
9. What is the difference between the posterior and posterior predictive distributions?
10. What is a hyper-hyper-parameter?
11. What is the key property of a conjugate prior?
12. In what setting is it unnecessary to include the q function in the Metropolis-Hastings acceptance probability?

Literature Survey

Reading academic papers is a skill that takes practice. When you first start out reading papers, you may find that you need to re-read things several times before you understand them, or that details will still be

¹The “positive” variables are some of the positive weights when you fit an L2-regularized logistic regression model to the this data. The “negative” variables are some of the negative regression weights in that model, and the “neutral” ones are set to 0 in that model.

very fuzzy even after you've put a great amount of effort into trying to understand a paper. Don't panic, this is normal.

Even if you are used to reading papers from your particular sub-area, it can be challenging to read papers about a completely different topic. Usually, people in different areas use different language/notation and focus on different issues. Nevertheless, many of the most-successful people in academia and industry are those that are able to understand/adapt ideas from different areas. (There are a ton of smart people in the world working on all sorts of amazing things, it's good to know how to communicate with as many of them as possible.)

A common technique when trying to understand a new topic (or reading scientific papers for the first time) is to read and write notes on 10 papers on the topic. When you read the first paper, you'll often find that it's hard to follow. This can make reading through it take a long time and might still leave you feeling that many things don't make sense; keep reading and trying to take notes. When you get to the second paper, it might still be very hard to follow. But when you start getting to the 8th or 9th paper, things often start making more sense. You'll start to form an impression of what the influential works in the area are, you'll start getting to used to the language and jargon, you'll start to understand what the main issues that people who work on the topic care about, and you'll probably notice some important references that weren't on your initial list of 10 papers. Ideally, you'll also start to notice how the topic has changed over time and you may get ideas of future work that you could do on the topic.

To help you make progress on your project or to give you an excuse to learn about a new topic, for this part you should [write a literature survey of at least 10 academic papers](#) on a particular topic. While your personal notes on the papers may be longer, the survey should be [at most 4 pages of text \(excluding references/tables/figures\)](#) in a format similar to the one for this document. Some logical components of a literature survey might be:

- A description of the overall topic, and the key themes/trends across the papers.
- A short high-level description of what was explored in each paper. For example, describe the problem being addressed, the key components of the proposed solution, and how it was evaluated. In addition, it is important to comment on the *why* questions: why is this problem important and why would this particular solution method make progress on it? It's also useful to comment on the strengths and weaknesses of the various works, and it's particularly nice if you can show how some works address the weaknesses of prior works (or introduce new weaknesses).
- One or more logical "groupings" of the papers. This could be in terms of the variant of the topic that they address, in terms of the solution techniques used, or in chronological terms.

Some advice on choosing the topic:

- The most logical/easy topic for your literature survey is a topic related to your course project, given that your final report will need a (shorter) literature survey included.
- If you are an undergrad, or a masters student without a research project yet, you may alternately want to choose a general area (like variance-reduced stochastic gradient, non-Gaussian graphical models, recurrent neural networks, matrix factorization, neural artistic style transfer, Bayesian optimization, transformer networks, etc.) as your topic.
- If you are a masters student that already has a thesis project, it could make sense to do a survey on a topic where ML intersects with your thesis (or where ML *could* intersect your thesis).
- If you are a PhD student, I would recommend using this an excuse to learn about a *completely different* topic than what you normally work on. Choose something hard that you would like to learn about, but previously haven't been able to justify spending the time exploring carefully. This can be invaluable to your future research, because during/after your PhD it often becomes hard to allocate time to learn completely new topics.