Math 189r
Homework 5
November 28, 2016

There are 3 problems in this set. Feel free to work with other students, but make sure you write up the homework and code on your own (no copying homework *or* code; no pair programming). Feel free to ask students or instructors for help debugging code or whatever else, though. When implementing algorithms you may not use any library (such as `sklearn`) that already implements the algorithms but you may use any other library for data cleaning and numeric purposes (`numpy` or `pandas`). Use common sense. Problems are in no specific order.

1 (Laplace Approximation) Reference Section 8.4 of Murphy on Bayesian Logstic Regression. We will use the Laplace Approximation to approximate the posterior distribution over $\mathbf{w}$ when we have a prior of the form $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_0)$. With the energy function $E(\mathbf{w}) = -\log \mathbb{P}(\mathcal{D}|\mathbf{w}) - \log \mathbb{P}(\mathbf{w})$,

(a) Compute the gradient of the energy ∇E, and

(b) Compute the Hessian of the energy $\nabla^2 E$.

(c) Using (a) and (b), what is the Laplace approximate posterior over $\mathbf{w}$? Assume we have the mode of the posterior $\mathbf{w}^\star$ such that $\nabla E(\mathbf{w}^\star) = 0$.

2 (Logistic Regression) Download the data at `https://math189r.github.io/hw/data/classification.csv`. Consider the Laplace Approximated Bayesian Logistic Regression from Problem 1. Calculate the posterior distribution over $\mathbf{w}$.

3 (Monte-Carlo Predictive Posterior) From Problem 2 we have the distribution $\mathbb{P}(\mathbf{w}|\mathcal{D})$. Now suppose we want to compute the probability that a test point $\mathbf{x}$ belongs to class 1. Analytically, we marginalize out $\mathbf{w}$ as

$$\mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathcal{D}) = \int \mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathbf{w})\mathbb{P}(\mathbf{w}|\mathcal{D})\, d\mathbf{w}.$$

Unfortunately, this integral cannot be computed in closed form (we say the integral is intractable). On the other hand, a Simple Monte Carlo approximation of the integral is

$$\mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathcal{D}) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathbf{w}^{(s)}). \qquad (\mathbf{w}^{(s)} \sim \mathbb{P}(\mathbf{w}|\mathcal{D}))$$

This is an unbiased estimate of the true predictive probability in the sense that its expectation is $\mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathcal{D})$. This is also easy to compute since we approximated $\mathbb{P}(\mathbf{w}|\mathcal{D})$ as a Gaussian, so we can sample from it easily.

(a) Given a function $f(\mathbf{x})$ where $\mathbf{x} \sim \mathbb{P}(\mathbf{x})$, show that

$$\mathbb{E}_{\mathbb{P}(\{\mathbf{x}^{(s)}\})}[\hat{f}] = \mathbb{E}\left[\frac{1}{S}\sum_{s=1}^{S} f(\mathbf{x}^{(s)})\right] = \mathbb{E}[f(\mathbf{x})]. \qquad (\mathbf{x}^{(s)} \sim \mathbb{P}(\mathbf{x}))$$

Put in other terms, show that our Monte Carlo estimator is unbiased.

(b) Show that the variance of the Monte Carlo estimate is proportional to $1/S$. That is, show

$$\mathbb{V}_{\mathbb{P}(\{\mathbf{x}^{(s)}\})}[\hat{f}] = \mathbb{V}[f(\mathbf{x})]/S.$$

Note that this means that standard deviation error bars shrink like $1/\sqrt{S}$.

(c) Plot the posterior predictive distribution $\mathbb{P}(\mathbf{y} = 1|\mathbf{x}, \mathcal{D})$ overlaying your data using this Monte Carlo approximation. You plot should look similar to Figure 8.6 in Murphy.