INF367: Spring 2020
Exercise 6

Instructions:

You can either return the solutions electronically via MittUiB by Monday 10.00 or show them on paper on Monday's meeting. Grades are awarded for effort so scanned notes are fine if you solve exercises by hand (no need to make fancy latex files).

Students are encouraged to write computer programs to derive solutions whenever appropriate.

Tasks

1. **Posterior of regression weights.**

Suppose $y_i = \mathbf{w}^T\mathbf{x}_i + \epsilon_i$, for $i = 1, \ldots, n$, where $\epsilon_i \sim \mathcal{N}(0, \beta^{-1})$. Assume a prior

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \alpha^{-1}\mathbf{I}).$$

Use 'completing the square' to show that the posterior of $\mathbf{w}$ is given by $P(\mathbf{w} \mid \mathbf{y}, \mathbf{x}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}, \mathbf{S})$, where

$$\mathbf{S} = \left(\alpha\mathbf{I} + \beta\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1},$$

$$\mathbf{m} = \beta\mathbf{S}\sum_{i=1}^{n} y_i\mathbf{x}_i.$$

2. **Poisson regression with Laplace approximation.**

Poisson regression can be used to model count data. That is, the label y_i tells how many times some event occurs (y_i is a non-negative integer). In this exercise, we try to predict the number of cyclists crossing Brooklyn Bridge given that we know precipitation in New York. A Poisson regression model can be defined as

$$y_i \mid \boldsymbol{\theta} \sim \text{Poisson}(\exp(\boldsymbol{\theta}^T\mathbf{x}_i)) \tag{1}$$

$$\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \alpha^{-1}\mathbf{I}) \tag{2}$$

where y_i are the observed counts, $\mathbf{x}_i$ the related covariates, $i = 1, \ldots, N$, and $\boldsymbol{\theta}$ are the regression weights. In this exercise, we approximate the posterior $P(\boldsymbol{\theta}|\mathbf{y})$ using the Laplace approximation.

(a) Derive the gradient $\nabla \log P(\boldsymbol{\theta}|\mathbf{y})$ and the Hessian $\mathbf{H} = \nabla\nabla \log P(\boldsymbol{\theta}|\mathbf{y})$ needed for the Laplace approximation.

Note that the posterior may look overly complicated but things get easier once you take the logarithm.

(b) Load data `new_york_bicycles.csv`. The data points are daily observations from year 2017. The data set has two variables. The first variable (x) is the daily precipitation in New York. The second variable (y) is the number of cyclists crossing the Brooklyn Bridge (measured in hundreds of cyclists).

(c) Split the data into training and test sets.

(d) Approximate the posterior distribution for parameters $\boldsymbol{\theta}$ given the training data using Laplace approximation. To get reasonable results, you need to use two-dimensional $\boldsymbol{\theta}$, that is, you should include an intercept term.

(e) Plot the true posterior and the approximation. Plot the difference between the true and approximate posterior in a third figure. Is the approximation reasonable?

Hint: Generate a grid for the parameter values (for example, using numpy's `meshgrid`). Compute unnormalized density on each grid point. Normalize by dividing the unnormalized densities by the sum over the whole grid. (This is not an exact normalizaton but it is close enough for our purposes.)

(f) Estimate credible intervals for the predicted number of cyclists (Note: the interval depends on x). Plot test data with precipitation on x-axis and the number of cyclists on y-axis. Plot the mean of the predictive distribution as a function of x. Plot the 50% credible intervals.

(g) Is your model well-calibrated? Does about 50% of test points lie within 50% credible interval (and similarly for 90% interval)?