EECS E6892 Topics in Information Processing
Bayesian Models for Machine Learning

Columbia University, Spring 2014

Homework 1

Due February 13. Show all work for full credit. Accepted one day late with 50% penalty.

Problem 1. Your friend is on a gameshow and phones you for advice. She describes her situation as follows: There are three doors with a prize behind one of the doors and nothing behind the other two. She randomly picks one of the doors, but before opening it, the gameshow host opens one of the other two doors to show that it contains no prize. She wants to know whether she should stay with her original selection or switch doors. What is your suggestion? Calculate the relevant posterior probabilities to convince her that she should follow your advice.

Problem 2. Let $\pi = (\pi_1, \ldots, \pi_K)$, with $\pi_j \geq 0, \sum_j \pi_j = 1$. Let $X_i \sim \text{Multinomial}(\pi)$, i.i.d. for $i = 1, \ldots, N$. Find a conjugate prior for π and calculate its posterior distribution and identify it by name. What is the most obvious feature about the parameters of this posterior distribution? Calculate $\mathbb{E}_p[\ln \pi_j]$ where the expectation is with respect to the posterior of π.

(Hint: The exponential family representation will be useful for the last part.)

Problem 3. You are given $\{x_1, \ldots, x_N\}$, where $x \in \mathbb{R}$. You model it as i.i.d. $\text{Normal}(\mu, \lambda^{-1})$. Since you don't know the mean μ or precision λ, you model them as $\mu|\lambda \sim \text{Normal}(0, a\lambda^{-1})$ and $\lambda \sim \text{Gamma}(b, c)$. Note that the priors are not independent. Using Bayes rule, calculate the posterior of μ and λ and identify the distributions.

Problem 4. Derive and implement a MAP inference algorithm for the multiclass logistic regression model. You can use the programming language of your choice. Using the MNIST digits data set provided on the course website, run your algorithm on the training set to learn regression vectors $w_0, \ldots, w_9$ for digits 0 through 9 respectively. Then, use these vectors to predict the labels of the data in the test set.

Show your results in a 10×10 confusion matrix using the predictive probabilities from the model (that is, don't map your prediction to the most probable class and then construct the confusion matrix). For example, if C is the confusion matrix, then after predicting the class of the ith digit as $Prob(class_i = k) = p_{i,k}$, update C as $C_{true\_class_i,k} = C_{true\_class_i,k} + p_{i,k}$ for each k.

Pick a testing example that the algorithm has a problem classifying. Show the reconstructed image according to the instructions in the readme file and give the probabilities of each class for that digit according to the model you learned.

Show your derivations and provide a print out of your code for full credit. If your code is long, please use a small font size to reduce the number of pages.