

**Homework 1: Due Friday, October 2, 2015 by 11:59pm****Please read these instructions to ensure you receive full credit on your homework.**

Submit the written portion of your homework as a *single* PDF file through Courseworks (less than 5MB). In addition to your PDF write-up, submit all code written by you in their original extensions through Courseworks (e.g., .m, .r, .py, etc.). Any coding language is acceptable. Do not wrap your files in .rar, .zip, .tar and do not submit your write-up in .doc or other file type. Your grade will be based on the contents of *one* PDF file and the original source code. Additional files will be ignored. We will not run your code, so everything you are asked to show should be put in the PDF file. Show all work for full credit.

**Late submission policy:** Late homeworks will have 0.1% deducted from the final grade for each minute late. *Your homework submission time will be based on the time of your **last** submission to Courseworks. I will not revert to an earlier submission!* Therefore, do not re-submit after midnight on the due date unless you are confident the new submission is significantly better to overcompensate for the points lost. Submission time is non-negotiable and will be based on the time you submitted your last file to Courseworks. The number of points deducted will be rounded to the nearest integer.

**Problem 1.** (10 points)

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.** (15 points)

Let  $\pi = (\pi_1, \dots, \pi_K)$ , with  $\pi_j \geq 0, \sum_j \pi_j = 1$ . Let  $X_i \sim \text{Multinomial}(\pi)$ , i.i.d. for  $i = 1, \dots, N$ . Find a conjugate prior for  $\pi$  and calculate its posterior distribution and identify it by name. What is the most obvious feature about the parameters of this posterior distribution?

**Problem 3.** (30 points)

You are given a dataset  $\{x_1, \dots, x_N\}$ , where each  $x \in \mathbb{R}$ . You model it as i.i.d.  $\text{Normal}(\mu, \lambda^{-1})$ . Since you don't know the mean  $\mu$  or precision  $\lambda$ , 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  $\mu$  and  $\lambda$  and identify the distributions.
- Using the posterior, calculate the predictive distribution on a new observation,

$$p(x^*|x_1, \dots, x_N) = \int_0^\infty \int_{-\infty}^\infty p(x^*|\mu, \lambda)p(\mu, \lambda|x_1, \dots, x_N)d\mu d\lambda$$

(Hint for 3b: Integrate out  $\mu$  first and then  $\lambda$  and take advantage of the factorization  $p(\mu, \lambda) = p(\mu|\lambda)p(\lambda)$ . The math is non-trivial, but these are common probability calculations, so feel free to search for the relevant distribution properties online, but make sure your derivation is clear. Also look for places where an integral corresponds to the normalizing constant of a distribution.)

**Problem 4.** (20 points)

In this problem you will use your derivations from Problem 3 to code a naive Bayes classifier for distinguishing handwritten 4's from 9's. The data for this problem and a description of it can be found on the course website and on Courseworks.

Each 15-dimensional vector  $x$  has a label  $y$  with  $y = 0$  indicating “4” and  $y = 1$  indicating “9”. We model the  $n$ th feature vector of a digit from class 1 as

$$p(x_n|\vec{\mu}_1, \vec{\lambda}_1, y_n = 1) = \prod_{d=1}^{15} \text{Normal}(x_{n,d}|\mu_{1,d}, \lambda_{1,d}^{-1}).$$

and similarly for class 0. We model the labels as  $y_n \sim \text{Bernoulli}(\pi)$ . Assume independent normal-gamma priors on all  $(\mu_{1,d}, \lambda_{1,d})$  and  $(\mu_{0,d}, \lambda_{0,d})$ , as in Problem 3, with  $a = 1$ ,  $b = 1$  and  $c = 1$ . For the label bias assume the prior  $\pi \sim \text{Beta}(e, f)$ .

Let  $(x^*, y^*)$  be a new test pair. The goal is to predict  $y^*$  given  $x^*$ . To do this we use the predictive distribution under the posterior of the naive Bayes classifier. That is, for possible label  $y^* = y \in \{0, 1\}$  we compute

$$p(y^* = y|x^*, X, \vec{y}) \propto p(x^*|y^* = y, \{x_i : y_i = y\})p(y^* = y|\vec{y})$$

where  $X$  and  $\vec{y}$  contain  $N$  training pairs of the form  $(x_i, y_i)$ . This can be calculated as follows:

$$p(x^*|y^* = y, \{x_i : y_i = y\}) = \prod_{d=1}^{15} \int_0^\infty \int_{-\infty}^\infty p(x^*|\mu_{y,d}, \lambda_{y,d})p(\mu_{y,d}, \lambda_{y,d}|\{x_i : y_i = y\})d\mu d\lambda$$

The results from Problem 3 can be directly applied here. Also

$$p(y^* = y|\vec{y}) = \int_0^1 p(y^* = y|\pi)p(\pi|\vec{y})d\pi$$

which has the solutions  $p(y^* = 1|\vec{y}) = \frac{e + \sum_n \mathbb{1}(y_n = 1)}{N + e + f}$  and  $p(y^* = 0|\vec{y}) = \frac{f + \sum_n \mathbb{1}(y_n = 0)}{N + e + f}$ .

- Using the marginal distributions discussed above, implement this naive Bayes classifier for binary classification in your preferred language.
- Make predictions for all data in the testing set by assigning the most probable label to each feature vector. In a  $2 \times 2$  table, list the total number of 4's classified as 4's, 9's classified as 9's, 4's classified as 9's, and 9's classified as 4's (i.e., a confusion matrix). Use the provided ground truth for this evaluation.
- Pick three misclassified digits and reconstruct the images as described in the readme file. Show these three images and their predictive probabilities.
- Pick the three most ambiguous predictions, i.e., the digits whose predictive probabilities are the closest to 0.5. Reconstruct the three images as described in the readme file and show them and their predictive probabilities.