# 277B: Machine Learning Algorithms

## Homework assignment #4: Statistical Models for Regression and Classification

## Assigned Feb. 20 and Due Mar. 1

**1. Baye's Theorem.** A chemical test is designed to indirectly determine whether the individual has a genetic marker predisposing him/her to having kidney disease. The chemical test has the following characteristics: the probability that a randomly chosen person who has the marker (M) will test positive (i.e. "marker present") is: P[+|M] = 0.95; the probability that a randomly chosen person who does not have the marker will test negative is: P[-|not M] = 0.95; the proportion of people who have the marker is: P[M] = 0.01.
(a) Define the following quantities: P[-|M]; P[+|not M]; P[not M].
(b) You have had a chemical test and have tested positive; should you be alarmed? To answer this, find what is the chance that a randomly selected person who tests positive for the marker actually has the marker by using Baye's Theorem. What feature of the given data accounts for the result?
(c) Suppose that frequency of marker was higher by a factor of 10, i.e. **P**[**M**] = 0.10. What is the chance that a randomly selected individual from this group who test positive actually has the marker?

**2. Gaussian Naïve Bayes.** We wish to classify 178 wines into 3 cultivars by solving it with Naïve Bayes. To do this we will classify the wines by assigning them to the cultivar with the largest P(cultivar | X), and to find this we must first define a labelled data set of P(wine attribute x | cultivar) pairings to learn the relationship where x is one of the attributes, and do this for all attributes.
(a)(2pt) How should we represent P(wine attribute x | cultivar) ? Fill in the code for Gaussian() function and give a reason that you choose this functional form. Given a wine that belongs to cultivar 1, what is the chance of it having an Alcohol % of 13 according to the probability distribution function?
(b)(4pt) Using your normalized chemical descriptor data, divide your data into 3-fold training and testing groups, i.e. using 2/3 training and 1/3 testing for the three divisions and report the results.

**3. Softmax and Cross Entropy Loss.** Work on the same wine dataset. Now we use another approach to do the classification. Implement a neural network model using PyTorch with no hidden layer (This is equivalent to a linear regression plus nonlinear activation function). Use softmax activation function in the last layer and use cross entropy loss as your loss function.
(a) (3pt) Pass the data through the network once without backpropagation and print out the output. Observe the difference between with and without the softmax activation layer. What does softmax do?
(b)(4pt) Divide data into 3-fold training and testing groups, within each fold further divide your training data into 80% training and 20% validation, choose the model for the epoch with lowest validation error. Report error in terms of success rate of classification. How good is the prediction relative to problem 2?