Homework 2

Due Date: April 9, 2016

This homework includes both math and coding problems. For submission, please refer to **SubmissionGuide.pdf**.

Problem 1. In classification, the loss function we usually want to minimize is the 0/1 loss:

$$l(f(x), y) = \mathbf{1}\{f(x) \neq y\}$$

where $f(x), y \epsilon \{0, 1\}$ (i.e., binary classification). In this problem we will consider the effect of using an asymmetric loss function:

$$l_{\alpha,\beta}(f(x), y) = \alpha \mathbf{1}\{f(x) = 1, y = 0\} + \beta \mathbf{1}\{f(x) = 0, y = 1\}$$

Under this loss function, the two types of errors receive different weights, determined by $\alpha, \beta > 0$.
Determine the Bayes optimal classifier, i.e. the classifier that achieves minimum risk assuming $\mathrm{P}(x, y)$ is known, for the loss $l_{\alpha,\beta}$ where $\alpha, \beta > 0$.

Problem 2. In this problem we will find the maximum likelihood estimator (MLE) and maximum a posteriori (MAP) estimator for the mean of a univariate normal distribution. Specifically, we assume we have N samples, $x_1, \cdots, x_N$ independently drawn from a normal distribution with *known* variance σ^2 and *unknown* mean μ.

1. Please derive the MLE estimator for the mean μ.
2. Now derive the MAP estimator for the mean μ. Assume that the prior distribution for the mean is itself a normal distribution with mean ν and variance β^2.
   HINT: $\beta^2 \left(\Sigma_{i=1}^N (x_i - \mu)^2 \right) + \sigma^2 (\mu - \nu)^2 =$
   $$\left[\mu \sqrt{N\beta^2 + \sigma^2} - \frac{\sigma^2 \nu + \beta^2 \Sigma_{i=1}^N x_i}{\sqrt{N\beta^2 + \sigma^2}} \right]^2 - \frac{\left[\sigma^2 \nu + \beta^2 \Sigma_{i=1}^N x_i \right]^2}{N\beta^2 + \sigma^2} + \beta^2 \left(\Sigma_{i=1}^N x_i^2 \right) + \sigma^2 \nu^2$$
3. Please comment on what happens to the MLE and MAP estimators as the number of samples N goes to infinity.

Problem 3. Consider the following training set, in which each example has two tertiary attributes (0, 1, or 2) and one of two possible classes (X or Y).

Example	A1	A2	Class
1	0	1	X
2	2	1	X
3	1	1	X
4	0	2	X
5	1	2	Y
6	2	0	Y

Please answer the following questions with details.

1. What would the Naive Bayes algorithm (with no smoothing) predict for the class of the following new example?

Example	A1	A2	Class
7	2	2	?

2. Apply Laplace smoothing (with a smoothing parameter equal to 1) to all the probabilities from the training set. Will the result of the previous question change?

Problem 4. Coding. Build a simple spam email classifier based on Nave Bayes.

1. Requirement:
   (a) Train the model using data in train.csv with all attributes provided and classify the email data in test.csv.
   (b) Output the classification accuracy ($\frac{the\ number\ of\ correctly\ classified\ emails}{the\ total\ number\ of\ emails}$)
2. Things should be included in your report:
   (a) A number indicating the classification accuracy
   (b) Tell me if you can achieve a higher classification accuracy using only some of the attributes provided.