COM S 474: Homework 2 Spring 2017

Write your name on each page. Maximum score is 30 points, due date is **Friday, February 17, 2017** . Please hand in the solutions (CLEAN version) on Friday, February 17, 2017 in class (hard copy). **Staple all the pages together**. NO credit will be given if work is not shown!! A good advice: if you don't know the answer, make an appointment with the TA or instructor. Do not simply google and copy/paste the answer. That is not a good practice and will not help you during the exam!

1. [15 points] Consider the "Magic Gamma Telescope" data set (you can download it from BlackBoard or it is freely available at `http://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope`). The data are MC generated to simulate registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. Cherenkov gamma telescope observes high energy gamma rays, taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas, and developing in the atmosphere. This Cherenkov radiation (of visible to UV wavelengths) leaks through the atmosphere and gets recorded in the detector, allowing reconstruction of the shower parameters. The data set consists out of 10 variables and 2 classes (last column of the data set). Find a suitable classifier that predicts gamma particles (signal) indicated as class -1 and hadron (background) indicated as class 1. Take the following training/test split: 13000 training and 6020 test. Perform all necessary tests, plots and Monte Carlo simulations to determine your final choice of classifier. You can use the built-in functions of your favorite program. Can you beat the 86.6% mean accuracy on test data (based on 100 runs)?

2. Consider the following data set

feature 1	feature 2	feature 3	class
0	0	0	0
1	0	1	1
1	0	0	0
1	1	1	1
0	1	1	1
1	1	0	0

   (a) [5 points] Classify the test point (0,0,1) using QDA and calculate the posterior class probabilities. Do the calculations by hand.

   (b) [5 points] Classify the test point (0,0,1) using naive Bayes assuming normality and calculate the posterior class probabilities. Do the calculations by hand. Use the estimators for the mean and the variance per class on p. 92 of the book or use the ones in the notes. These are the same as for one dimensional LDA.

   (c) [5 points] Verify your result for both classifiers using Matlab, R, etc. You can use the built-in functions.