Assignment 03

To be solved individually

Submit by 18 Nov, 2018, 23h59 by email to jaime.cardoso@fe.up.pt

1. Load the height/weight data from the file heightWeightData.txt. The first column is the class label (1=male, 2=female), the second column is height, the third weight.

Start by replacing the weight column by the product of height and weight.

For the Fisher's linear discriminant analysis as discussed in the class, send the python/matlab code and answers for the following questions:

a. What's the S_B matrix?

b. What's the S_W matrix?

c. What's the optimal 1d projection direction?

d. Project the data in the optimal 1d projection direction. Set the decision threshold as the middle point between the projected means. What's the misclassification error rate?

e. What's your height and weight? What's the model prediction for your case (male/female)?

2. Consider the Logistic Regression as discussed in the class. Assume now that the cost of erring an observation from class 1 is cost1 and the cost of erring observations from class 0 is cost0. How would you modify the goal function, gradient and hessian matrix (slides 11 and 12 in week 5)?

Change the code provided (or developed by you) in the class to receive as input the vector of costs. Test your code with the following script:

```
trainC1 = mvnrnd([21 21], [1 0; 0 1], 1000);
trainC0 = mvnrnd([23 23], [1 0; 0 1], 20);
testC1 = mvnrnd([21 21], [1 0; 0 1], 1000);
testC0 = mvnrnd([23 23], [1 0; 0 1], 1000);
NA = size(trainC1,1);
NB = size(trainC0,1);
traindata = [trainC1 ones(NA,1); trainC2 zeros(NB,1)]; %add class label in the last column
weights=logReg(traindata(:,1:end-1),traindata(:,end),[NB NA])
testC1 = [ones(size(testC1,1),1) testC1]; %add virtual feature for offset
testC0 = [ones(size(testC0,1),1) testC0]; %add virtual feature for offset
%FINISH the script to compute the recall, precision and F1 score in the test data
```

In this script the cost of erring in C1 is proportional to the elements in C0. Compute the precision, recall and F1 in the test data. **Note: if you are unable to modify to account for costs, solve without costs.**

3. Several phenomena and concepts in real life applications are represented by angular data or, as is referred in the literature, directional data. Assume the directional variables are encoded as a periodic value in the range $[0, 2\pi]$.
Assume a two-class (y_0 and y_1), one dimensional classification task over a directional variable x, with equal a priori class probabilities.

a) If the class-conditional densities are defined as $p(x|y_0)= e^{2\cos(x-1)}/(2\ \pi\ 2.2796)$ and $p(x|y_1)= e^{3\cos(x+0.9)}/(2\ \pi\ 4.8808)$, what's the decision at x=0?
b) If the class-conditional densities are defined as $p(x|y_0)= e^{2\cos(x-1)}/(2\ \pi\ 2.2796)$ and $p(x|y_1)= e^{3\cos(x-1)}/(2\ \pi\ 4.8808)$, for what values of x is the prediction equal to y_0?
c) Assume the more generic class-conditional densities defined as $p(x|y_0)= e^{k_0\cos(x- \mu_0)}/(2\ \pi\ I(k_0))$ and $p(x|y_1)= e^{k_1\cos(x-\mu_1)}/(2\ \pi\ I(k_1))$. In these expressions, k_i and μ_i are constants and $I(k_i)$ is a constant that depends on k_i. Show that the posterior probability $p(y_0|x)$ can be written as $p(y_0|x) = 1/(1+e^{w_0+ w_1\sin(x- \Theta)})$, where w_0, w_1 and Θ are parameters of the model (and depend on k_i , μ_i and $I(k_i)$).