**B429/529 Assignment 1**

For Q1, please 1) manually calculate the values step by step on the physical paper, as the examples in the slides, and 2) Write the python codes according to the manual calculation step by step.

**Question 1.** Use the perceptron algorithm to manually find a weight vector **w** = [*w*0,*w*1,*w*2]T  for linear classification of 3 data points (**x***i*,*yi*) satisfying sign(**w**T**x***i*) = *yi*. The input data are

**x**1=(1,2,0)T *y*1=+1

**x**2=(1,2,2)T *y*2=+1

**x**3=(1,0,2)T *y*3=-1

The first element in vector **x***i* is always *x*0=1 (See slides in lecture 1).

In the first step of the perceptron algorithm, the initial assignment of **w** is [0,6,6]T. The intermediate steps of the computation are needed. (25 points)

**Question 2.** (a) Use Python to compute the weight vector **w** = [*w*0,*w*1,*w*2]T for linear regression of *N* input data points (**x***i*,*yi*) that minimizes the sum of squared error (**w**T**x***i* –*yi*)2 (**use the method in slide 29 in lecture 2**). The input data are

**x**1=(1,3,2)T *y*1 = 3

**x**2=(1,2,2)T *y*2 = 1

**x**3=(1,1,-1)T *y*3 = -1

**x**4=(1,-1,-3)T *y*4= -3

(b) Use the linearRegression module in Python package sklearn to find **w** for linear regression. Steps of the python code and intermediate results are needed (25 points).

**Question 3.** In Fisher’s linear discriminant analysis, we search for a vector **w** such that all data points are well separately after they are projected to the direction defined by **w**. The input data points are

**x**1=(4,3)T *y*1 = 1

**x**2=(2,2)T *y*2 = 1

**x**3=(1,1)T *y*3 = -1

**x**4=(-2,-1)T *y*4= -1

(Note that the first element in vector **x***i* is NOT *x*0=1). Use Python to find a vector **w**=[*w*1,*w*2] with Fisher’s linear discriminant analysis. Steps of the python code and intermediate results are needed (**use the steps in slide 43 in lecture 2**) (25 points).

**Question 4.** A heart disease data set (file heart\_disease\_data.txt in the folder) contains the medical information of 100 patients of two classes. The first row contains column names, and other rows contain the information of the patients. The last column “num” contains the classes of the patients. Use the perceptron module in Python package sklearn to find the weight vector **w** for linear classification (15 points). Use the LDA module in Python package sklearn to find the weight vector **w** for linear discriminant analysis. The python code of the analysis should be provided (10 points).