**B529 Assignment 1**

**Due date: 6pm, Feb 10, 2015**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Question 1.** Please 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 is

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

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

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

The first element in vector **x** 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. Please provide the intermediate steps of the computation. (30 points)

**Question 2.** Please use R to manually 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. The input data is

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

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

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

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

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

Please provide the intermediate steps of the computation (20 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**. When the input data points are

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

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

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

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

(Note that the first element in vector **x** is NOT *x*0=1 in this problem), please use R to manually find a vector **w**=[*w*1,*w*2] with Fisher’s LDA and provide the intermediate steps (20 points).

**Question 4**. In the problem of determining if a peptide-spectrum-match is a correct one or not, we have the following problem:

* Input: scores of peptide spectrum matches: **x***i* = [*x*]
* Output: if the spectrum is generated from the peptide: *yi* = +1/-1

Prior probity: P(*y* = 1) = 0.1 P(*y* = -1) = 0.9

Likelihood: (N(10, 4)) and

(N(5, 1))

Cost function

Correct prediction

Output function

*h*(**x**)

|  |  |  |
| --- | --- | --- |
| *e***(***f***(x),** *g***(x))** | +1 | -1 |
| +1 | 0 | 1000 |
| -1 | 1 | 0 |

Please compute the classification function that minimizes the cost (error). Please provide the intermediate steps of the computation. (30 points)