

Zhiyuan(James) Zhang

CS 513 Midterm

October 25, 2017

I pledge my honor that I have abided by the Stevens Honor System.

-----James

1A $100/400 = 0.25$

1B $149/400 = 0.3725$

1C $46 / 400 = 0.115$

1D $34 / 400 = 0.085$

1E $15 / 80 = 0.1875$

1F No, let's do the Hispanic cocaine addiction cell.

The fraction of $7 / 46$ (cocaine given Hispanic) is not the same the $149 / 400$ (pure cocaine without given any race).

0.15 does not equal to 0.3725 .

2 see excel

3 Property 1: Distance is always non-negative

Property 2: Commutative, distance from "A to B" is distance from "B to A"

Property 3: Triangle inequality holds, distance from "A to C" must be less than or equal to distance from "A to B to C"

It breaks the third rule. This function fulfills the property 1 and 2, but it doesn't work on because if we do the difference separately, then $(0,0)$ to $(0,1)$ plus $(0,1)$ to $(1,1)$ is smaller than directly doing $(0,0)$ to $(1,1)$. It should not be smaller than it, so this is not a proper distance function.

Distance from $(0,0)$ to $(0,1)$ is 1

Distance from $(0,1)$ to $(1,1)$ is 1

Distance from $(0,0)$ to $(1,1)$, however, according to the given equation, would be 8. It is clearly not the correct answer.

4a see excel

4b see excel

5 We simply delete all the record with the question marks to prevent any unnecessary errors.

For details please see the R program.

Comments are all in the R problem, and the comments explains everything.

The following is the screen shot of our test result:

```

> training<-readfile[-myData,]
> knn5<-knn(training[,-10], test[,-10], training[,10], k=5)
> knn5
 [1] 2 4 2 2 4 2 2 4 4 2 4 4 4 4 2 2 2 4 2 2 2 4 4 4 2 4 2 2 4 2 4
[33] 4 2 2 2 4 4 4 4 4 4 2 4 4 4 4 2 4 4 2 4 4 4 4 4 4 2 4 2 2 4
[65] 4 4 4 4 2 4 2 2 2 2 2 2 2 2 2 4 2 4 2 4 4 2 2 2 2 4 2 2 2 2
[97] 2 2 2 2 4 2 2 4 2 2 2 2 2 2 4 4 2 2 2 4 2 2 4 4 2 2 4 2 4 2 2
[129] 2 2 2 2 2 4 2 2 4
Levels: 2 4
> err5<-1-sum(test[,10]==knn5)/length(knn5)
> err5
[1] 0.05109489
> table(Prediction=knn5,Actual=test[,10] )
      Actual
Prediction 2  4
      2 74  4
      4  3 56
> |

```