Junhao Zhang “Freddie”

USC ID: 1943-0782-95

CSCI 567 – HW2

09/14/2021

*jzhang49@usc.edu*

**[Problem 1]**

1. **The answer is ▲**. Because it is the majority among the 4 nearest neighbors.
2. If *K* = *N*, then covers the entire set. **The answer is ▲** because it is the majority among the entire set.
3. **The answer is 2**. Because when *K* = *1*, everyone is classified to what its nearest neighbor is. Thus, as shown in the screenshot below only these two would be classified correctly.

Chart

Description automatically generated

**[Problem 2]**

**(*2.1*)**

Since *M* is a positive define matrix, thus there exists some matrix *A* such that . Then we observe that:

= =

Thus, the original equation becomes:

=

whose Optimization Objective is:

*RSS* (*w*) = + , which is the *m*-th row of *A*.

And we are going to find stationary points:

= + 2

=

Let *0* = , then we have:

*0* =

=

=

=

**=**

**(*2.2*)**

* # 1

We need to find the maximum likelihood estimation for of the product of for all *n*. In other words:

For finding the maximum likelihood estimation, we need to take the natural log instead:

So, we have:

And we take the partial derivative with respect to the variable , and get:

Then, we set it to equal to 0 and get the following:

* # 2

**is the same as above**. And we take the partial derivative with respect to the variable

σ, and get:

Then, we set it to equal to 0 and get the following:

**[Problem 3]**

**(*3.1*)**

based on the definition of , we had proven if the algorithm makes a mistake, the update rule moves the weight *w* towards the direction of the optimal weights *.*

**(*3.2*)**

Next, we need to cancel from both sides of the inequation and try to prove:

So, we do the following:

, and we can commutate to .

But , so . Therefore, we can remove it from the inequation and try to prove:

So, we do the following:

, and we can commutate to .

For the same reason above, we can remove it from the inequation and try to prove:

Because the problem description states that , which further indicates that , and , thus can be at most 1. By this step, we had already proven .

**(*3.3*)**

Let’s assume the algorithm makes mistake at every single iteration. Because is true, so the followings are also true:

,

,

•

•

•

Moreover, because and , so . Moreover, because and , so .

•

•

•

So, we can see each time the algorithm makes a mistake, the constant behind will increment. Therefore, if totally made *M* mistakes in the first *k* iterations, we can see:

And the problem description states that , thus:

which further indicates that:

Again, we assume the algorithm makes mistake at every single iteration. Because is true, so the followings are also true:

,

,

•

•

•

,

Moreover, because and , so:

Moreover, because and , so:

•

•

•

So, we can see each time the algorithm makes a mistake, the constant increments. Therefore, if totally made *M* mistakes in the first *k* iterations, we can see:

And the problem description states that , thus:

And because for any two vectors *a* and *b*, so , which further indicates that:

And the problem description also states that , thus:

We had proven that .

**(*3.4*)**

Because , so , so:

, which in another form: