**INT 305 Assignment 1**

(The deadline is 31st of Oct.)

1. Please write down the whole derivation process to obtain the gradient for logistic regression.

(30%)

Solution:

The logistic model is: .  
and the activation function is: ,  
the loss function is:

To optimize the model, we should update the weight by using gradient descent.

Therefore, by chain rule:

1), because ,

2), because ,

3), because ,

Therefore,

Now, we can get the gradient of , then we need to update weight .

2. Please   write   down  the   whole   derivation   process   to   obtain   the   gradient  for  multiclass

classification with softmax. (40%)

Solution:

The model function is: ,

and the softmax function is:

and the cross-entropy loss function is:

To optimize the model, we should update the weight by using gradient descent.

Therefore, by chain rule:

1), because ,

Since after one-hot encoding, for classification task with classes, is a -dimensional vector. And only the position of the correct class is 1, and the rest is 0.

So, there are two cases: for correct class , , . However, according to the function , we only need to consider the condition of , because the rest condition will get 0. For the same reason, we only need to consider as well.

Therefore, ,

so,

2), because ,

Since mentioned before, there are also two cases.

i), For :

because is one of ,

so,

Thus,

ii), For :

3), because ,

4), because we have two cases of , so we should consider them both when calculate .

i), For ,

ii), For ,

Therefore, the result of are both .

5),

Now, we can get the gradient of , then we need to update weight .

3. Please  compare  the  SVM  loss  and  Softmax  loss  for  multiclass  classification,  please  explain

which one is better? (30%)

Solution:

The softmax loss is better for multiclass classification.

For example, now there is a multiclass classification task with three classes. And we sample three training examples, their scores are shown below (the bold type class is label):

|  |  |  |  |
| --- | --- | --- | --- |
|  | Class 1 | Class 2 | Class 3 |
| Example 1 | **10** | -2 | 3 |
| Example 2 | **10** | 9 | 9 |
| Example 3 | **10** | -100 | -100 |

Now, we calculate SVM loss and Softmax loss for these three examples respectively.

The formula of softmax loss is:

The formula of SVM loss is:

The results are shown below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Class 1 | Class 2 | Class 3 | SVM | Softmax |
| Example 1 | **10** | -2 | 3 | 0 | 0.4E-3 |
| Example 2 | **10** | 9 | 9 | 0 | 0.24 |
| Example 3 | **10** | -100 | -100 | 0 | 0 |

It can be seen from the results that SVM loss cannot reflect the degree of model optimization precisely. When using SVM Loss to optimize the model, we may only find the local optimal solution, because after obtaining a solution, SVM Loss will become 0.

However, Softmax loss doesn't have that problem, it is a good reflection of the current model. Moreover, even if we find a solution, we can continue to optimize until we find the optimal solution.

Therefore, the Softmax loss is better than SVM loss for multiclass  classification.