|  |
| --- |
| **PA2 Individual Report** |

**Zhenrui Yue**

Computer Science & Engineering

UC San Diego

La Jolla, CA 92093

*yuezrhb@gmail.com*

**1 Maximum Likelihood Estimation**

Given the probability function:

To find the maximum likelihood estimate for the parameter for n samples, we introduce the log likelihood :

To maximize is the same as to minimize :

We differentiate with respect to :

Let the first equation above equal 0, we have:

Since the second derivative is greater than 0, the value above proves to be the minimum of , therefore this is the value for maximum likelihood estimation.

**2 Multiclass Classification**

1. **Derivation**

Given the cross-entropy loss and softmax function below:

We first compute the derivative of the softmax function with respect to :

We also compute the derivative of the softmax function with respect to :

Now we compute the derivative of the cross-entropy loss with respect to :

Plug in the first 2 computed derivatives, we have:

Therefore, for the output layer we have (note that notation for output layer is k):

For the hidden layer we first compute the derivative of :

So, we have for the hidden layer, note represents the class of true label:

There for the result for the hidden layer is:

1. **Update Rule**

For the output layer with n data samples, note was already computed in (a), and represents the output of the class with true label:

Whereas for the input layer, could also be plugged in:

1. **Vectorize computation**

For matrix computation we have following equations, with representing the one-hot coded label matrix, notice that it’s matrix multiplication between matrix and matrix (vector) but element-wise computation with functions:

The backpropagation would be in the following order, with representing the calculated probability of the class with the correct label,