CpE 520: HW#5

West Virginia University

Ali Zafari

Contents

Notation **2**

Question 1 **3**

Binary Logistic Regression - Squared Error 3

Question 2 **5**

Binary Logistic Regression - Cross Entropy 5

Notation

In this homework assignment we will use superscripts in parentheses to denote the sample number and subscripts to denote the feature number.

$x_i^{(j)}$	:	value of i'th feature of the j'th input sample
$y^{(j)}(w)$	:	predicted label of the j'th sample (which is a function of wieghts i.e. w_i's)
$t^{(j)}$	:	real label of the j'th sample

Question 1

Binary Logistic Regression - Squared Error

The gradient descent updating rule for each weight in neural network is

$$w_i^{t+1} = w_i^t - \eta \frac{\partial}{\partial w_i} J(w) \tag{1}$$

in which $J(w)$ is the cost function and η is learning rate.

In this question our cost function is Squared Error, as given below

$$J(w) = \frac{1}{2} \sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w))^2$$

and we are going to minimize this cost by choosing the best values for $w_1, ..., w_n$ with the help of gradient descent updating rule i.e. equation 1.
Therefore the first thing to be done is to find $\frac{\partial}{\partial w_i} J(w)$:

$$\begin{aligned}\frac{\partial}{\partial w_i} J(w) &= \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w))^2 \\ &= \frac{1}{2} \sum_{j=1}^{N} -2(t^{(j)} - y^{(j)}(w)) \frac{\partial y^{(j)}(w)}{\partial w_i} \\ &= -\sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w)) \frac{\partial y^{(j)}(w)}{\partial w_i} \end{aligned} \tag{2}$$

So now, we need to know $\frac{\partial y^{(j)}(w)}{\partial w_i}$.
$y^{(j)}(w)$ is repeated below to use them for derivation:

$$\begin{cases} y^{(j)}(w) = \frac{1}{1+exp(-z^{(j)}(w))} \\ z^{(j)}(w) = \sum_{i=1}^{n} w_i x_i^{(j)} \end{cases}$$

and after gettingderivatives of both equations, we have:

$$\begin{cases} \frac{\partial y^{(j)}(w)}{\partial w_i} = \frac{e^{-z^{(j)}}}{(1+e^{-z^{(j)}})^2} \frac{\partial z^{(j)}(w)}{\partial w_i} = \frac{1}{(1+e^{-z^{(j)}})} \frac{e^{-z^{(j)}}}{(1+e^{-z^{(j)}})} \frac{\partial z^{(j)}(w)}{\partial w_i} = y^{(j)}(w)(1 - y^{(j)}(w)) \frac{\partial z^{(j)}(w)}{\partial w_i} \\ \frac{\partial z^{(j)}(w)}{\partial w_i} = x_i^{(j)} \end{cases}$$

so we now have $\frac{\partial y^{(j)}(w)}{\partial w_i}$ as:

$$\frac{\partial y^{(j)}(w)}{\partial w_i} = y^{(j)}(w)(1 - y^{(j)}(w)) x_i^{(j)} \tag{*}$$

The last thing to do is to replace the above eqution into equation 2:

$$\frac{\partial}{\partial w_i} J(w) = -\sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w)) y^{(j)}(w)(1 - y^{(j)}(w)) x_i^{(j)}$$

Updating rule of gradient descent(equation 1) will be:

$$\boxed{w_i^{t+1} = w_i^t + \eta \sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w)) y^{(j)}(w)(1 - y^{(j)}(w)) x_i^{(j)}}$$

Question 2

Binary Logistic Regression - Cross Entropy

We follow the same procedure as we did in question 1, except for the definition of cost function. The new cost function is defined below:

$$J(w) = -\sum_{j=1}^{N} t^{(j)} \log y^{(j)}(w) + (1 - t^{(j)}) \log\Big(1 - y^{(j)}(w)\Big)$$

we find its derivative as:

$$\begin{aligned}\frac{\partial}{\partial w_i} J(w) &= \frac{\partial}{\partial w_i}[-\sum_{j=1}^{N} t^{(j)} \log y^{(j)}(w) + (1 - t^{(j)}) \log\Big(1 - y^{(j)}(w)\Big)] \\ &= -\sum_{j=1}^{N} t^{(j)} \frac{\frac{\partial y^{(j)}(w)}{\partial w_i}}{y^{(j)}(w)} - (1 - t^{(j)}) \frac{\frac{\partial y^{(j)}(w)}{\partial w_i}}{1 - y^{(j)}(w)} \qquad (3)\end{aligned}$$

we now replace the $\frac{\partial y^{(j)}(w)}{\partial w_i}$ in the numerators of 3 with the result of equation (*) from previous question as shown below:

$$\begin{aligned}w_i^{t+1} &= w_i^t + \eta \sum_{j=1}^{N} t^{(j)} \frac{y^{(j)}(w)(1 - y^{(j)}(w))x_i^{(j)}}{y^{(j)}(w)} - (1 - t^{(j)}) \frac{y^{(j)}(w)(1 - y^{(j)}(w))x_i^{(j)}}{1 - y^{(j)}(w)} \\ &= w_i^t + \eta \sum_{j=1}^{N} x_i^{(j)} [t^{(j)} - t^{(j)} y^{(j)}(w) - y^{(j)}(w) + t^{(j)} y^{(j)}(w)]\end{aligned}$$

$$\boxed{w_i^{t+1} = w_i^t + \eta \sum_{j=1}^{N} (t^{(j)} - y^{(j)}(w)) x_i^{(j)}}$$