AI6102 Assignment

Name: *BAIWEN*

Matric number: *G1903363K*

Question 1

For $c = 0$, $P(y=0|x) = p_c = \frac{1}{1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}$, which means that $W^{(0)} = 0$

For $c > 0$, $P(y=c|x) = p_c = \frac{\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}{1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}$

Using Cross entropy Loss Function

$$L = -\sum_j y_j \log p_j$$

As in multiple class problem, only the right class $j = c$ has $y_j = 1$, the rest $y_j = 0|j \neq c$, so the Loss function converted into

$$L = -\log p_c$$

$$\frac{\partial L}{\partial W^{(i)}} = -\frac{1}{p_c}\frac{\partial p_c}{\partial W^{(i)}}$$

For $c = 0, i \neq c$,

$$\begin{aligned}
\frac{\partial L}{\partial W^{(i)}} &= -\frac{}{p_c}\frac{\partial p_c}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{\partial \frac{1}{1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{-(\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))'}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}\frac{\exp(-W^{(i)T}X)X}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}p_c p_i X \\
&= -p_i X
\end{aligned}$$

For $c = 0, i = c$,

$$\frac{\partial L}{\partial W^{(i)}} = 0$$

For $c > 0, i \neq c$,

$$
\begin{aligned}
\frac{\partial L}{\partial W^{(i)}} &= -\frac{1}{p_c}\frac{\partial p_c}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{\partial \frac{\exp(-W^{(c)T}X)}{1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{(\exp(-W^{(c)T}X))'(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)) - (\exp(-W^{(c)T}X))(\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))'}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}\frac{(\exp(-W^{(c)T}X))(\exp(-W^{(i)T}X))}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}p_c p_i X \\
&= -p_i X
\end{aligned}
$$

For $c > 0, i = c$,

$$
\begin{aligned}
\frac{\partial L}{\partial W^{(i)}} &= -\frac{1}{p_c}\frac{\partial p_c}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{\partial \frac{\exp(-W^{(c)T}X)}{1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)}}{\partial W^{(i)}} \\
&= -\frac{1}{p_c}\frac{(\exp(-W^{(c)T}X))'(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X)) - (\exp(-W^{(c)T}X))(\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))'}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}\frac{-(\exp(-W^{(c)T}X))(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))X + (\exp(-W^{(c)T}X))(\exp(-W^{(c)T}X))}{(1+\sum_{c=1}^{C-1}\exp(-W^{(c)T}X))^2} \\
&= -\frac{1}{p_c}(-p_c + {p_c}^2) \\
&= (1-p_c)X
\end{aligned}
$$

Then we can substitute the above result into the following equation to get the update rule for the $W^{(i)}$:

$$
W_{t+1}^{(i)} = W_{t+1}^{(i)} - \lambda\frac{\partial L}{\partial W^{(i)}}
$$

Question 2

Table 1: The 3-fold cross-validation results of varying values of C in LinearSVC on the a5a training set (in accuracy).

	C = 0.01	**C=0.1**	**C=1**	**C=10**	**c=100**
Accuracy of linear SVMs	0.845546	0.848872	0.848612	0.847728	0.778089

C=0.1 is the best value

Table 2: Test results of LinearSVC with the best value of C on the a5a test set (in accuracy).

	C = 0.1
Accuracy of linear SVMs	0.850639