CSC311 HW3

Yesinn Kwak 1004603734

March 21st, 2020

Question 1

(a)

The dimension of x is $[d, 1]$: dimensional row vector. From $(z_2 = h + x)$ we can deduce that, z_2, h, x all three has the same dimension therefore the dimension of z_2, h is also $[d, 1]$. From $h = \sigma(z_1)$, as σ is an activation function, it does not have any affect on dimension, dimension of $\sigma(z_1)$ is the same as z_1, which is the same as h. Therefore the dimension of z_1 is also $[d, 1]$. From $z_1 = W^{(1)}x$, since we know the dimension of z_1 and x, $W^{(1)}x$ is a matrix dot product, whose result is of dimension z_1 $[d, 1]$. Let the dimension of $W^{(1)}$ be $[d1, d2]$. Since $[d, 1] = [d1, d2] \cdot [d, 1]$ we can say that $d1 = d$ and $d2 = d$. Therefore the dimension of $W^{(1)}$ is $[d, d]$. From: $L = \frac{1}{2}(y - t)^2$ with $t \in \mathbf{R}$ we know t is a real number therefore y also has to be a real number. Dimension of y and t must be $[1, 1]$. We also know that from $y = W^{(2)}z_2$ we know that $W^{(2)}z_2$ is a matrix dot product. Let the dimension of $W^{(2)}$ be $[d3, d4]$, $[1, 1] = [d3, d4] \cdot [d, 1]$. From this we can conclude that $d3 = 1$ and $d4 = d$, Therefore the dimension of $W^{(2)}$ is $[1, d]$. In conclusion the Dimensions are:

- $W^{(1)}$: $[d, d]$
- $W^{(2)}$: $[1, d]$
- z_1: $[d, 1]$
- z_2: $[d, 1]$

(b)

We need to know the number of elements in $W^{(1)}$ and $W^{(2)}$ to calculate the number of parameters in the network. The size of $W^{(1)}$ is $[d, d]$ therefore there are d x $d = d^2$ parameters in $W^{(1)}$. The size of $W^{(2)}$ is $[1, d]$, so there are 1 x $d = d$ parameters in $W^{(2)}$. The total parameters in the network as a function of d is: parameters in $W^{(1)}$ + parameters in $W^{(2)} = d^2 + d$

(c)

- With respect to y:

$$\begin{aligned}\frac{\partial L}{\partial y} &= \frac{\partial}{\partial y}(\frac{1}{2}(y-t)^2) \\ &= \frac{1}{2}\frac{\partial}{\partial y}(y-t)^2 \\ &= \frac{1}{2} \times 2(y-t)\frac{\partial y}{\partial y} \\ &= (y-t) \times 1 \\ &= (y-t)\end{aligned}$$

- With respect to $W^{(2)}$:

$$\begin{aligned}\frac{\partial L}{\partial W^{(2)}} &= \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial W^{(2)}} \\ \frac{\partial L}{\partial y} &= (y-t) \\ \frac{\partial y}{\partial W^{(2)}} &= \frac{\partial}{\partial W^{(2)}}(W^{(2)} z_2) \\ &= z_2 \times \frac{\partial W^{(2)}}{\partial W^{(2)}} \\ &= z_2 \\ \frac{\partial L}{\partial W^{(2)}} &= z_2(y-t)\end{aligned}$$

- With respect to z_2:

$$\begin{aligned}\frac{\partial L}{\partial z_2} &= \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \\ \frac{\partial y}{\partial z_2} &= \frac{\partial}{\partial z_2}(W^{(2)} z_2) \\ &= W^{(2)} \\ \frac{\partial L}{\partial z_2} &= W^{(2)}(y-t)\end{aligned}$$

- With respect to h:

$$\begin{aligned}\frac{\partial L}{\partial h} &= \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} \\ \frac{\partial z_2}{\partial h} &= \frac{\partial}{\partial h}(h+x) = 1 \\ \frac{\partial L}{\partial h} &= (y-t) \times W^{(2)} \times 1 = W^{(2)}(y-t)\end{aligned}$$

- With respect to z_1:

$$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} \times \frac{\partial h}{\partial z_1}$$

$$\frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} = W^{(2)}(y-t)$$

$$\frac{\partial h}{\partial z_1} = \frac{\partial}{\partial z_1}(\sigma(z_1))$$

$$= \sigma'(z_1)$$

$$\frac{\partial L}{\partial z_1} = \sigma'(z_1)W^{(2)}(y-t)$$

- With respect to $W^{(1)}$:

$$\frac{\partial L}{\partial W^{(1)}} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} \times \frac{\partial h}{\partial z_1} \times \frac{\partial z_1}{\partial W^{(1)}}$$

$$\frac{\partial z_1}{\partial W^{(1)}} = \frac{\partial}{\partial W^{(1)}}(W^{(1)}x)$$

$$= x$$

$$\frac{\partial L}{\partial W^{(1)}} = \sigma'(z_1)W^{(2)}(y-t)x$$

- With respect to x:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} \times \frac{\partial h}{\partial z_1} \times \frac{\partial z_1}{\partial x} + \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial x}$$

$$\frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} \times \frac{\partial z_2}{\partial h} \times \frac{\partial h}{\partial z_1} = \frac{\partial L}{\partial z_1} = \sigma'(z_1)W^{(2)}(y-t)$$

$$\frac{\partial L}{\partial y} \times \frac{\partial y}{\partial z_2} = \frac{\partial L}{\partial z_2} = W^{(2)}(y-t)$$

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z_1} \times \frac{\partial z_1}{\partial x} + \frac{\partial L}{\partial z_2} \times \frac{\partial z_2}{\partial x}$$

$$= \sigma'(z_1)W^{(2)}(y-t) \times \frac{\partial z_1}{\partial x} + W^{(2)}(y-t) \times \frac{\partial z_2}{\partial x}$$

$$\frac{\partial z_1}{\partial x} = \frac{\partial}{\partial x}(W^{(1)}) = W^{(1)}$$

$$\frac{\partial z_2}{\partial x} = \frac{\partial}{\partial x}(h+x) = 1$$

$$\frac{\partial L}{\partial x} = W^{(2)}(y-t)\left[\sigma'(z_1)W^{(1)} + 1\right]$$

Question 2

(a)

- When $k = k'$

$$\begin{aligned}
\frac{\partial y_k}{\partial z_{k'}} &= \frac{\partial}{\partial z_{k'}} \\
&= \frac{e^{z_{k'}}}{\sum_{j=1}^{K} e^{z_j}} \\
&= \frac{e^{z_{k'}}}{\sum_{j=1}^{K} e^{z_j}} - \frac{e^{z_{k'}}}{\left(\sum_{j=1}^{K} e^{z_j}\right)^2} e^{z_{k'}} \\
&= y_k - y_k^2
\end{aligned}$$

- When $k \neq k'$

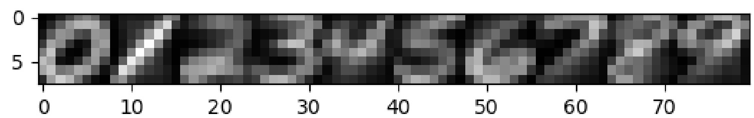
$$\begin{aligned}
\frac{\partial y_k}{\partial z_{k'}} &= \frac{\partial}{\partial z_{k'}} \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \\
&= -\frac{e^{z_k}}{\left(\sum_{j=1}^{K} e^{z_j}\right)^2} e^{z_{k'}} \\
&= -y_k y_{k'}
\end{aligned}$$

(b)

$$\begin{aligned}
\frac{\partial L_{CE}(t, y(x;W))}{\partial w_k} &= \frac{\partial L_{CE}}{\partial z_k} \frac{\partial z_k}{\partial w_k} \\
\frac{\partial L_{CE}}{\partial z_k} &= -\sum_k t \frac{\partial \log y}{\partial z_k} = -\sum_k t \frac{1}{y} \frac{\partial y}{\partial z_k} \\
&= -t_k(1 - y_k) - \sum_{k \neq i} t \frac{1}{y}(-y_k y) \\
&= -t_k(1 - y_k) + \sum_{k \neq i} t(y_k) \\
&= -t_k + t_k y_k + \sum_{k \neq i} t(y_k) \\
&= y_k \left(\sum_k t\right) - t_k = y_k - t_k \\
\frac{\partial z_k}{\partial w_k} &= x \\
\frac{\partial L_{CE}(t, y(x;W))}{\partial w_k} &= \frac{\partial L_{CE}}{\partial z_k} \frac{\partial z_k}{\partial w_k} = (y_k - t_k)x
\end{aligned}$$

Question 3

(0)



3.1 K-NN Classifier

(1)

(a) For $K = 1$ the training set classification accuracy is 1.000000 and the testing set classification accuracy is 0.968750
(b) For $K = 15$ the training set classification accuracy is 0.963714 and the testing set classification accuracy is 0.960750

(2)

In the situation of a tie, you can decrease the size of K until the tie is broken. Since the K-NN measures how close the target is related to the trained model, decreasing the value of K counts the points that are much more closely related to the target data. This will guarantee that the tie is broken and won't affect the accuracy of the method.

(3)

Using K= 1 classifier accuracy is: 0.964429
Using K= 2 classifier accuracy is: 0.964429
Using K= 3 classifier accuracy is: 0.965143
Using K= 4 classifier accuracy is: 0.965571
Using K= 5 classifier accuracy is: 0.9634294
Using K= 6 classifier accuracy is: 0.964571
Using K= 7 classifier accuracy is: 0.960714
Using K= 8 classifier accuracy is: 0.961571
Using K= 9 classifier accuracy is: 0.958000
Using K= 10 classifier accuracy is: 0.956857
Using K= 11 classifier accuracy is: 0.955571
Using K= 12 classifier accuracy is: 0.955000
Using K= 13 classifier accuracy is: 0.953143
Using K= 14 classifier accuracy is: 0.954286
Among all the K-NN, the classifier performed the best when K is 4. When K = 4, the training set classification accuracy is 0.986429, the testing set classification accuracy is 0.972750 and the average accuracy across folds is 0.965571

3.2 Classifiers Comparison

see q3_2.py

3.3 Model Comparison

The accuracy of the SVM on the test set is: 0.93225
The accuracy of ADABoost with the SAMME algorithm is: 0.8945
The SVM performed better than the ADABoost but the K-NN model outperformed both of the models.