Stamatios Morellas

COM S 474 - Final Exam

11/24/2020

Pledge

> "I affirm that the work on this exam is my own and I will not use any people to help me nor will I share any part of this exam or my work with others without permission of the instructor." – Stamatios Morellas

Question 1

The result of the Hadamard product $A \circ B$ is shown below:

$$A \circ B = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1/2 & 1 & 6 \\ 3 & -4 & 2 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & 2 \\ 1 & -2 & 2 \end{pmatrix}$$

Question 2

The result of AB^T:

$$AB^T = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1/2 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 3 \\ 1 & -4 \\ 6 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1.66667 \\ 6.66667 & 1 \end{pmatrix}$$

The result of BA^T:

$$BA^T = \begin{pmatrix} 0.5 & 1 & 6 \\ 3 & -4 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1/3 \\ 1/2 & 1/2 \\ 1/3 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 6.66667 \\ 1.66667 & 1 \end{pmatrix}$$

Question 3

Given $f(x) = x + 1$, the value of $f(AB^T) = \mathbf{AB^T + 1}$.

Now, we compute $f(AB^T)$:

$$f(AB^T) = \begin{pmatrix} 4 & 2.66667 \\ 7.66667 & 2 \end{pmatrix}$$

Question 4

Given that $d = 3$, $\mathbf{x} = [x_0, x_1, x_2, x_3] = [1/2, 1/3, 1/4, 1/5]^T$, and $\mathbf{w} = [w_0, w_1, w_2, w_3] = [2, 3, 4, 5]$, and $\phi(x) = x^2$, we can find $\hat{y}$ by following the steps below.

First, we must find $w^T x$:

$$w^T x = \begin{pmatrix} 2 & 3 & 4 & 5 \end{pmatrix} * \begin{pmatrix} 1/2 \\ 1/3 \\ 1/4 \\ 1/5 \end{pmatrix} = 4$$

Then, we use $\phi(x) = x^2$ to find $\phi(w^T x)$:

$$\phi(w^T x) = \phi(4) = 4^2 = 16$$

So from this, we get:

$$\hat{y} = \phi(w^T x) = (w^T x)^2 = \mathbf{16}$$

Question 5

a. The value of $\frac{\partial E}{\partial \hat{y}} = \mathbf{1}$.

b. The value of $\frac{\partial \hat{y}}{\partial (w^T x)} = 2 * w^T x = 2 * 4 = \mathbf{8}$.

c. The value of $\frac{\partial (w^T x)}{\partial x_1} = \frac{\partial (w_0 x_0 + w_1 x_1 + w_2 x_2 + w_3 x_3)}{\partial x_1} = w_1 = \mathbf{3}$.

d. The value of $\frac{\partial E}{\partial x_1} = (\frac{\partial E}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial (w^T x)} * \frac{\partial (w^T x)}{x_1}) = 1 * 8 * 3 = \mathbf{24}$.

Question 6

We saw from question 5 that $\frac{\partial E}{\partial x_1} = 1 * 8 * w_1 = \mathbf{8w_1}$ and $\frac{\partial E}{\partial w_1} = 1 * 8 * x_1 = \mathbf{8x_1}$.

Now, to find the values of $\frac{\partial E}{\partial x}$ and $\frac{\partial E}{\partial w}$, we will use the following equations to find the results:

For $\frac{\partial E}{\partial \mathbf{x}}$:

$$\frac{\partial E}{\partial \mathbf{x}} = 8 * \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 16 \\ 24 \\ 32 \\ 40 \end{pmatrix}$$

For $\frac{\partial E}{\partial \mathbf{w}}$:

$$\frac{\partial E}{\partial \mathbf{w}} = 8 * \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/3 \\ 1/4 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 4 \\ 8/3 \\ 2 \\ 8/5 \end{pmatrix}$$

Question 7

The values for all activations $x^{(l)}$ for all $l \in [1..3]$ are as follows:

We start with the transfer matrices for every pair of layers:

For $l = 0$:

$$\left(\mathbb{W}^{(0)}\right)^T * \mathbf{x}^{(0)} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 0.3 \end{pmatrix}$$

$$\mathbf{x_{[1..]}}^{(1)} = \phi(\left(\mathbb{W}^{(0)}\right)^T * \mathbf{x}^{(0)}) = \sigma \begin{pmatrix} 3 \\ -3 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0.3 \end{pmatrix} \rightarrow \mathbf{x}^{(1)} = \begin{pmatrix} 1 \\ 3 \\ 0 \\ 0.3 \end{pmatrix}$$

For $l = 1$:

$$\left(\mathbb{W}^{(1)}\right)^T * \mathbf{x}^{(1)} = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 0 \\ 0.3 \end{pmatrix} = \begin{pmatrix} 2.15 \\ 2.15 \end{pmatrix}$$

$$\mathbf{x_{[1..]}}^{(2)} = \phi((\mathbb{W}^{(1)})^T * \mathbf{x}^{(1)}) = \sigma \begin{pmatrix} 2.15 \\ 2.15 \end{pmatrix} = \begin{pmatrix} 2.15 \\ 2.15 \end{pmatrix} \rightarrow \mathbf{x}^{(2)} = \begin{pmatrix} 1 \\ 2.15 \\ 2.15 \end{pmatrix}$$

For $l = 2$:

$$\left(\mathbb{W}^{(2)}\right)^T * \mathbf{x}^{(2)} = \begin{pmatrix} 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 \end{pmatrix} \begin{pmatrix} 1 \\ 2.15 \\ 2.15 \end{pmatrix} = \begin{pmatrix} 1.325 \\ 1.325 \end{pmatrix}$$

$$\mathbf{x}^{(3)} = \phi((\mathbb{W}^{(2)})^T * \mathbf{x}^{(2)}) = \sigma \begin{pmatrix} 1.325 \\ 1.325 \end{pmatrix} \rightarrow \mathbf{x}^{(3)} = \begin{pmatrix} 1.325 \\ 1.325 \end{pmatrix}$$

Question 8

If the loss is squared error, $E = (\hat{y} - y)^2$, we find $\delta^{(3)}$ by doing the following:

$$\delta^{(3)} = (\hat{y} - y)^2 = \begin{pmatrix} (1.325 - y_1)^2 \\ (1.325 - y_2)^2 \end{pmatrix}$$

where y is $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

Since the ψ function derived for ReLU activation is equal to 1, we have the following:

$$\delta^{(2)} = \psi(x^{(2)}) \circ (\mathbb{W}^{(2)}\delta^{(3)}) = \mathbb{W}^{(2)}\delta^{(3)}$$

or

$$\delta^{(2)} = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} * \delta^{(3)}$$

Then, we find $\delta^{(1)}$ as follows:

$$\delta^{(1)} = \psi(x^{(1)}) \circ (\mathbb{W}^{(1)}\delta^{(2)}_{[1..]}) = \mathbb{W}^{(1)}\delta^{(2)}_{[1..]}$$

or

$$\delta^{(1)} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} * \delta^{(2)}_{[1..]}$$

Finally, we find $\delta^{(0)}$:

$$\delta^{(0)} = \psi(x^{(0)}) \circ (\mathbb{W}^{(0)}\delta^{(1)}_{[1..]}) = \mathbb{W}^{(0)}\delta^{(1)}_{[1..]}$$

or

$$\delta^{(0)} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix} * \delta^{(1)}_{[1..]}$$

Question 9

A model with regularization of parameters prevents the network from overfitting the data, thus resulting in better generalization.