Uralp Ergin 5975013, İsmail Karabaş 7654321, Ahmet Yaşar Ayfer 5986167

DL EX 6

1) $$\begin{bmatrix}1\\2\\3\\4\end{bmatrix} \; \begin{bmatrix}2\\1\\3\end{bmatrix} \longrightarrow \begin{bmatrix}2\cdot 1 + 1\cdot 2 + 3\cdot 3 + \overset{b}{1} \\ 2\cdot 2 + 1\cdot 3 + 3\cdot 4 + 1\end{bmatrix} = \begin{bmatrix}14\\20\end{bmatrix}$$

$$y - \hat{y} = \begin{bmatrix}2\\4\end{bmatrix} - \begin{bmatrix}14\\20\end{bmatrix} = \begin{bmatrix}-12\\-16\end{bmatrix}$$

$$L = \frac{1}{2}\|y - \hat{y}\|_2^2 = \frac{1}{2}\left((-12)^2 + (-16)^2\right) = \frac{1}{2}\cdot 400$$

$$\underline{L = 200}$$

2) $$\frac{\partial L}{\partial \hat{y}_i}\left(\frac{\partial \hat{y}_i}{\partial b}\right) = \underline{-28}$$

$$\frac{\partial \hat{y}_i}{\partial b} \to 1$$

$$\frac{\partial L}{\partial \hat{y}_i} = \sum_i y_i - \hat{y}_i = -12 + (-16) = \underline{-28}$$

$$\frac{\partial L}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial w_1} = \overset{-12}{(y_1 - \hat{y}_1)}\, x_1 + \overset{-16}{(y_2 - \hat{y}_2)}\, x_2$$

$$= -12 \cdot 1 + (-16)\cdot 2 = -44$$

$$\frac{\partial L}{\partial w_2} = -12 \cdot x_2 + (-16)\cdot x_3 = -12 \cdot 2 + (-16)\cdot 3$$

$$= -72$$

$$\frac{\partial L}{\partial w_3} = -12\, x_3 + (-16)\, x_4 = -12 \cdot 3 + (-16)\cdot 4$$

$$= -36 - 64 = -100$$

$$\frac{\partial L}{\partial w} = \begin{bmatrix}-44\\-72\\-100\end{bmatrix}$$

3) $w' = w - \alpha \frac{dL}{dw}$

$$= \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - 0.01 \begin{bmatrix} -44 \\ -72 \\ -100 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} -0.44 \\ -0.72 \\ -1 \end{bmatrix} = \begin{bmatrix} 2.44 \\ 1.72 \\ 4 \end{bmatrix}$$

$$b' = b - 0.01 \cdot \underbrace{\frac{\partial L}{\partial b}}_{-28} = 1 + 0.28$$

$$= \underline{\underline{1.28}}$$

<u>New Forward Pass:</u>

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2.44 \\ 1.72 \\ 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1.(2.44) + 2.(1.72) + 3 \cdot 4 + b \\ 2 \cdot (2.44) + 3(1.72) + 4 \cdot 4 + b \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} 19.16 \\ 27.32 \end{bmatrix}$$

$$L = \frac{1}{2} \|y - \hat{y}\|_2^2 = \frac{1}{2} \cdot \left[(-17.16)^2 + (-23.32)^2 \right]$$

$$\|y - \hat{y}\| = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 19.16 \\ 27.32 \end{bmatrix} = \begin{bmatrix} -17.16 \\ -23.32 \end{bmatrix}$$

$$\frac{1}{2} \cdot (838.14) = \underline{\underline{419.07}}$$

Experiments

* In both MLP and convolution accuracy results dropped on shifted data but convolution outperformed MLP on both normal and shifted data. It can be said that MLP is not able recognize and classify correctly when there is a change in input. It behaves like its a whole new input and because it only uses fully connected layers which lacks capturing the spatial relationships.

* We can use data augmentation for that. Shifts, rotations, flips, scaling can be applied to training data (increasing the variety in data) so that the model can learn the patterns not based on their positions. Or can use ensemble methods like combining a model trained on shifted data with a model trained on normal data so that the combined model can leverage both of the models.

* Because subsequent operations like pooling, finite receptive fields, training biases, and fully connected layers can disrupt this property and hence, CNNs may not preserve this property always as seen in the results.