Deeplearning HK6 Participants. 1° Hanzhi Zhuang 5578347 2° Rui Song 5576534 3 Tengyunhao Yang 5587382

1. Pen & Paper Tasks

1) $\hat{y}_i = \sum_j w_j x_{i+j} + b$.

$$\hat{y} = \begin{bmatrix} [2\,1\,3]\cdot[\frac{1}{3}]+1 \\ [2\,1\,3]\cdot[\frac{2}{4}]+1 \end{bmatrix} = \begin{bmatrix} 14 \\ 20 \end{bmatrix}$$

144+256.

$$L(y,\hat{y}) = \frac{1}{2}\|y-\hat{y}\|_2^2 = \frac{1}{2}\left\|\begin{bmatrix}2\\4\end{bmatrix}-\begin{bmatrix}14\\20\end{bmatrix}\right\|_2^2 = \frac{1}{2}\left\|\begin{bmatrix}-12\\-16\end{bmatrix}\right\|_2^2 = 200$$

2) $\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial w}$ $\quad \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial b}$

$\frac{\partial L}{\partial \hat{y}_1} = \hat{y}_1 - y_1 = 14-2 = 12$. $\quad \frac{\partial L}{\partial \hat{y}_2} = \hat{y}_2 - y_1 = 20-4 = 16$.

$\frac{\partial \hat{y}_i}{\partial w_i} = \frac{\partial(w_i x_i + b)}{\partial w_i} = x_i$ $\quad \frac{\partial \hat{y}_i}{\partial b} = \frac{\partial(w_i x_i + b)}{\partial b} = 1$.

$$\therefore \frac{\partial L}{\partial w} = \begin{bmatrix}\hat{y}_1 - y_1 & \hat{y}_2 - y_2\end{bmatrix}\cdot\begin{bmatrix}x_1 & x_2 & x_3\\ x_2 & x_3 & x_4\end{bmatrix} = \begin{bmatrix}44\\72\\100\end{bmatrix}$$

$$\frac{\partial L}{\partial b} = \sum \frac{\partial L}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial b} = 12+16 = 28$$

3) $\alpha = 0.01$.

$$W' = W - \alpha\cdot\frac{\partial L}{\partial w} = \begin{bmatrix}2\\1\\3\end{bmatrix} - 0.01\cdot\begin{bmatrix}44\\72\\100\end{bmatrix} = \begin{bmatrix}1.56\\0.28\\2\end{bmatrix}$$

$$b' = b - \alpha\cdot\frac{\partial L}{\partial b} = 1 - 0.01\cdot 28 = 0.72.$$

forward passing:

$$\hat{y}'_i = \sum_j W'_j x_{ij} + b'.$$

$$\hat{y}' = \begin{bmatrix} [1.56 \ 0.28 \ 2] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ [1.56 \ 0.28 \ 2] \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \end{bmatrix} + \begin{bmatrix} 0.72 \\ 0.72 \end{bmatrix}$$

$$= \begin{bmatrix} 8.12 \\ 11.96 \end{bmatrix} + \begin{bmatrix} 0.72 \\ 0.72 \end{bmatrix}$$

$$= \begin{bmatrix} 8.84 \\ 12.68 \end{bmatrix}.$$

$$L' = \frac{1}{2} \| y' - \hat{y}' \|_2^2 = \frac{1}{2} \left\| \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 8.84 \\ 12.68 \end{bmatrix} \right\|_2^2 = 61.064.$$

6. Equivariance.

2)

To do 1: The accuracies of CNN model are slightly higher than than those of MLP model both on the original and shifted dataset.

Because the CNN has translation equivalence feature for shifted dataset, also it can capture more spatial relationships and hierarchical features.

Todo2: Because MLP model are more likely to become overfitting, we can apply regularization method to make the model more generalized, such as augmentation of dataset, insert dropout and batchnormalization in the network.

Todo3:

Because CNN are equivariant to translation, so the location of the object in the image is important, if the input image is shifted or translated, the output will also be shifted accordingly.

The model is trained on unshifted data, so it may not recognize the same shifted pattern on the shifted dataset.