Submission - Assignment 6

Marta Gulida - matriculation number 5585808 - mg776
Erik Bode - matriculation number 4505199 - kb301
Tillman Heisner - matriculation number 4517815 - th273

Pen and Paper Task

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ A kernel $w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$, a scalar bias b, a stride of 1, no padding

Expected output $\hat{y} = \begin{bmatrix} \hat{y_1} \\ \hat{y_2} \end{bmatrix}$. The loss function $L(y, \hat{y}) = \frac{1}{2} \|y - \hat{y}\|_2^2$.

1) Perform a single forward pass of this cross-correlation with stride 1 and no padding. Then, plug in the following values for x, w and b to get $\hat{y}$:

$x = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ $w_0 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ $b_0 = 1$ $y = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ $y_i = \sum_j w_j x_{i+j}$

Stride 0: $\hat{y_1} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} [x_1\ x_2\ x_3] + b = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$

$\hat{y_1} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} [1\ 2\ 3] + 1 = 2 + 2 + 9 + 1 = 14$

Stride 1: $\hat{y_2} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} [x_2\ x_3\ x_4] + b = w_1 x_2 + w_2 x_3 + w_3 x_4 + b$

$\hat{y_2} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} [2\ 3\ 4] + 1 = 4 + 3 + 12 + 1 = 20 \Rightarrow \hat{y} = \begin{bmatrix} 14 \\ 20 \end{bmatrix}$

Loss: $L(y, \hat{y}) = \sum_{i=1}^{2} \frac{1}{2} \|y_i - \hat{y_i}\|_2^2 = \frac{1}{2} \sum_{i=1}^{2} \|y_i - \hat{y_i}\|_2^2$

$L(y, \hat{y}) = \frac{1}{2} ((2-14)^2 + (4-20)^2) = \frac{1}{2} \cdot 400 = 200$

2) Perform a single backward pass of this convolution, i.e., find $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$

$\frac{\partial L(\hat{y}, y)}{\partial \hat{y}} = \begin{bmatrix} \hat{y_1} - y_1 \\ \hat{y_2} - y_2 \end{bmatrix}$ $\frac{\partial L(\hat{y}, y)}{\partial \hat{y}} = \begin{bmatrix} 12 \\ 16 \end{bmatrix}$

$\frac{\partial L(\hat{y}, y)}{\partial w} = \sum_i \frac{\partial L(\hat{y}, y)}{\partial \hat{y_i}} \cdot \frac{\partial \hat{y_i}}{\partial w} = [\hat{y_1} - y_1] \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [\hat{y_2} - y_2] \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} =$

$= [14 - 2] \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + [20 - 4] \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 12 \\ 24 \\ 36 \end{bmatrix} + \begin{bmatrix} 32 \\ 48 \\ 64 \end{bmatrix} = \begin{bmatrix} 44 \\ 72 \\ 100 \end{bmatrix}$

$$\frac{\partial L(\hat{y},y)}{\partial b} = \sum_i \frac{\partial L(\hat{y},y)}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial b} = 12 + 16 = 28$$

This is 1 and we have summation

3) Now, plug in the values from (1) above to get the updated values of w and b at the end of the backward pass. Assume a learning rate of 0.01. Perform one more forward pass with the updated weights and bias. What is the loss now?

$$w_1 = w_0 + \alpha \cdot \frac{\partial L(\hat{y},y)}{\partial w}$$

$$w_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - 0.01 \begin{bmatrix} 44 \\ 72 \\ 100 \end{bmatrix} = \begin{bmatrix} 2 - 0{,}01 \cdot 44 \\ 1 - 0{,}01 \cdot 72 \\ 3 - 0{,}01 \cdot 100 \end{bmatrix} = \begin{bmatrix} 1{,}56 \\ 0{,}28 \\ 2 \end{bmatrix}$$

$$b_1 = b_0 - \alpha \cdot \frac{\partial L(\hat{y},y)}{\partial b}$$

$$b_1 = 1 - 0{,}01 \cdot 28 = 0{,}72$$

Forward pass with w_1 and b_1:

Stride 0: $$\hat{y}_1 = \begin{bmatrix} 1{,}56 \\ 0{,}28 \\ 2 \end{bmatrix} [1 \ 2 \ 3] + 0{,}72 = 1{,}56 \cdot 1 + 0{,}28 \cdot 2 + 2 \cdot 3 + 0{,}72 = 8$$

Stride 1: $$\hat{y}_2 = \begin{bmatrix} 1{,}56 \\ 0{,}28 \\ 2 \end{bmatrix} [2 \ 3 \ 4] + 0{,}72 = 1{,}56 \cdot 2 + 0{,}28 \cdot 3 + 2 \cdot 4 + 0{,}72 = 12{,}68$$

$$\hat{y} = \begin{bmatrix} 8{,}84 \\ 12{,}68 \end{bmatrix} \qquad L(\hat{y},y) = \frac{1}{2} \sum_{i=1}^{2} \| y_i - \hat{y}_i \|_2^2 = \frac{1}{2}\left((2 - 8{,}84)^2 + (4 - 12{,}68)^2 \right) =$$

$$= \frac{1}{2} \left(46{,}7856 + 75{,}3424 \right) = 61{,}064$$

Equivariance →

Equivariance

1. What do you observe regarding the accuracies? How can the results be explained?

Accuracy on data using an MLP: 0.9488
Accuracy on shifted data using an MLP: 0.4037
Accuracy on data using convolutional model: 0.9672
Accuracy on shifted data using convolutional model: 0.594

Given the results above, we observe an accuracy reduction of 57.45% for the MLP and 38.59% for the convolutional model. They can be explained by the architectural differences.
Because of the convolutions, the CNN is more invariant to translation than the MLP.
The moving kernel is equivariant to the translational changes.

2. What could we do to improve the accuracy of the MLP model on the shifted validation set?

To improve the accuracy of the MLP, it should be trained on an augmented dataset, including shifted images.
Enlarging the network for an increased learning capacity should also be viable.

3. Why is the performance of the CNN not equal for unshifted and shifted data, if convolutions are equivariant to translation?

The performance difference of the convolutional network should be in part caused by the fully connected layer.
Some parts of the network are equivariant: The translations of the input propagate through the network via the convolutional layers, shifting the feature maps.
The final fully connected layer is not equivariant, but the previous convolutions decreased the difference in the shifted set visible to it.
This results in better performance than the MLP, but not the same level as before.