1D Convolution Operation

Dr. Alireza Aghamohammadi

May 28, 2024

Understanding Convolutional Neural Networks (CNNs)

- Convolutional Neural Networks (CNNs) are composed of multiple convolutional layers, which are equivariant to translation.
- These networks often include pooling layers that provide partial invariance to translation.
- Convolutional layers perform operations based on the mathematical concept of *convolution*.
- For simplicity, this lecture will focus on one-dimensional (1D) convolutions.

1D Convolution in Neural Networks

- In 1D convolution, an input vector $\mathbf{x}$ is transformed into an output vector $\mathbf{z}$, where each output z_i is a weighted sum of nearby input values.
- The same set of weights, known as the *convolution kernel* or *filter*, is applied at each position.
- The *kernel size* refers to the number of input elements combined to produce each output element.

$$z_i = w_1 x_{i-1} + w_2 x_i + w_3 x_{i+1} \tag{1}$$

Illustration of 1D Convolution

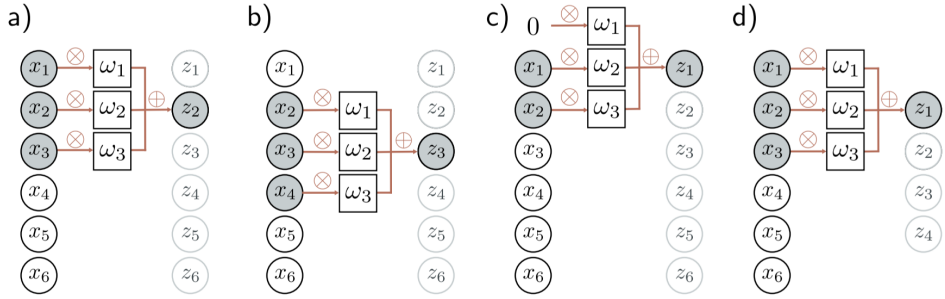

Figure: 1D convolution operation[1]

[1]Adapted from the book "Understanding Deep Learning"

Handling Boundaries in 1D Convolution

- The equation

$$z_i = w_1 x_{i-1} + w_2 x_i + w_3 x_{i+1}$$

  demonstrates that each output is computed as a weighted sum of the previous, current, and subsequent input positions.
- How do we handle the first and last output positions?
- With *zero padding*, we assume the input is zero outside its valid range.
- With *valid padding*, we discard output positions where the kernel exceeds the input range.