Appendix

A No representation ability lose for 1D-CNN

q is convolution kernel which receptive field is L. For any values in q, we could find a two-layer CNN of the same receptive filed and the network can extract same information as q. For instance, q is a convolution kernel which size is 5, and we have a two-layer network the kernel size of the first layer is 2 and the kernel size of second layer is 4. No matter what values in q, we could use the two-layer network get same result via adjusting values in each layer.

For simplicity the proof is based on the dilation=1 and the stride=1, and do not add Relu or Batch Norm in equations. Actually, BN and Relu will not influence the conclusion. It can be seen in the code verification here[13]. We add BN and Relu in the code, and the conclusion does not change. This is because the bias term in Batch Norm is trainable, thus it can reduce the influence of Relu.

The proof is below: let's assume the input signal in receptive field is X which length is L. The convolutions result of X and q is

$$\sum_{l=1}^{L} x_l q_l + b$$

Assume the length of kernel k is N and we have N channels in first layer. k_m^n denotes the m-th value of k in the n-th channel. when bias = 0 and

$$k_m^n = \begin{cases} 1, if \ m = n \\ 0, if \ m \neq n \end{cases}$$

the convolutional result of first layer is

$$\begin{bmatrix} x_1, x_2, ..., x_{L-N+1} \\ x_2, x_3, ..., x_{L-N+2} \\ ... \\ x_N, x_N + 1, x_3, ..., x_L \end{bmatrix}$$

If the input channel of w is N and out put channel is one, the convolutional result of the second layer is

$$sum \left(\begin{bmatrix} x_1, x_2, ..., x_{L-N+1} \\ x_2, x_3, ..., x_{L-N+2} \\ ... \\ x_N, x_{N+1}, ..., x_L \end{bmatrix} \cdot \begin{bmatrix} w_{1,1}, w_{1,2}; ..., w_{1,L-N+1} \\ w_{2,1}, w_{2,2}; ..., w_{2,L-N+2} \\ ... \\ w_{N,1}, w_{N,2}, ..., w_{N,L-N+1} \end{bmatrix} \right) + b \tag{9}$$

where Equation 9 can also be written as

$$\sum_{l=1}^{L} x_l \left(\sum_{i=1}^{l} w_{i,l+1-i} \right) + b \tag{10}$$

we could see that when $\sum_{i=1}^{l} w_{i,l+1-i} = q_l$ the convolutional result of second layer is same as the x and q. Thus, w and k exist.

[13] https://github.com/Anonymouslink/OS-CNN/blob/master/Code_example_of_theoretical_proof/4_3_Check_No_representation_ability_lose.ipynb