
# Homework3

**Course:** Machine Learning

**Name:** Xiang Lei (雷翔)

**Student ID:** 2053932

**Date:** June 2024

**Q1.** Briefly answer the following questions:

(1) Why do neural networks need activation functions?

(2) Briefly describe the influence of the value of the learning rate on the training of the neural network.

(3) What advantages does CNN have over fully connected DNN in image classification?

**Ans.** (1) If there are no activation functions in neural networks, each layer of the neural network would simply perform a linear transformation on the input, resulting in the entire network being equivalent to a single linear transformation regardless of the number of layers. If activation functions are used in neural networks, the neural networks can fit nonlinear functions, enabling them to learn and represent intricate features and relationships in the input data.

(2) If the value of the learning rate on the training of the neural network is too low, requiring many iterations to converge to a minimum, which can lead to higher computational costs and time. If it is too high, the model may overshoot the optimal solution, causing the training process to become unstable and potentially diverge.

(3) Compared with fully connected DNNs, CNNs exploit the spatial hierarchy of features in images by using convolutional layers, which preserve the spatial relationships between pixels. The parameters of CNNs are generally much less than fully connected DNNs. The feature extraction ability of CNNs is more powerful than fully connected DNNs in image classification.

**Q2.** The input size of AlexNet is 227×227×3. Given CONV1:

(CONV1): 96 11x11 filters at stride 4, pad 0

What is the output size after CONV1? Write out the calculation process.

**Ans.**

$$W_{out} = \frac{W_{in} - K + 2P}{S} + 1 = \frac{227 - 11 + 2 \times 0}{4} + 1 = 55$$

$$H_{out} = \frac{H_{in} - K + 2P}{S} + 1 = \frac{227 - 11 + 2 \times 0}{4} + 1 = 55$$

So the output size after CONV1 is 55×55×96.

**Q3.** Convolution and Pooling Operations：

(1) The 4×4 feature map is convolved with a 3×3 convolution kernel (stride = 1).

Calculate feature maps output in both cases respectively:

a) No padding;

b) 0 is padded around the feature map to keep the output feature size unchanged

Feature map:

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \end{bmatrix}$$

Kernel

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix}$$

(2) The given feature maps are max-pooled and average-pooled respectively using a 2×2 pooling layer (stride = 2). Calculate the output feature map.

Feature map:

$$\begin{bmatrix} 1 & 4 & 2 & 1 \\ 5 & 8 & 3 & 4 \\ 7 & 6 & 4 & 5 \\ 1 & 3 & 1 & 2 \end{bmatrix}$$

**Ans.** (1)

(a) No padding:

$$\begin{bmatrix} 15 & 16 \\ 6 & 15 \end{bmatrix}$$

(b) keep the output size unchanged

$$W_{out} = \frac{W_{in} - K + 2P}{S} + 1 = \frac{4 - 3 + 2 \times P}{1} + 1 = 4$$

So P = 1.

Feature map:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 0 \\ 0 & 3 & 0 & 1 & 2 & 0 \\ 0 & 2 & 3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The output:

$$\begin{bmatrix} 7 & 12 & 10 & 2 \\ 4 & 15 & 16 & 10 \\ 10 & 6 & 15 & 6 \\ 8 & 10 & 4 & 3 \end{bmatrix}$$

(2) The output of max-pooled:

$$\begin{bmatrix} 8 & 4 \\ 7 & 5 \end{bmatrix}$$

The output of average-pooled:

$$\begin{bmatrix} 4.5 & 2.5 \\ 4.25 & 3.0 \end{bmatrix}$$