Homework 3

ID	Name
1953902	GAO Yangfan

Table of Contents

Problem 1

(1)

As most problems of classification is not linear, non-linear activation functions are needed to better fit non-linear models by adding non-linear factors.

(2)

Just like what learning rate affects linear & logistic regression, there are 2 situations:

- **Learning rate is too small**: this will cause a very long time of training as it descents too slow.
- **Learning rate is too big**: this will cause huge steps of descent which may not converge while training.

(3)

1. The number of parameter of CNN is much less than that in a fully connect DNN as CNN shares the weights

2. CNN is capable of detecting certain features everywhere in the image but DNN is only able to detect the features in certain regions

Problem 2

As the size of output image can be calculated by:

$$\frac{N-F}{\text{stride}}+1$$

where:

N is size of imput image (with padding)

F is size of convolution kernel

as the pad here is 0, $N=227$, $F=11$

The size of output image is:

$$\frac{227-11}{4}+1=55$$

And since there are 96 filters, the output size is $55\times 55\times 96$

Problem 3

(1)

a

The output size is:

$$\frac{4-3}{1}+1=2$$

Denote the feature map as I and the convolutional kernal as K

The $(1,1)$ element of output can be computed by:

$$\sum_{i=1}^{3}\sum_{j=1}^{3} I_{ij}K_{ij} = 15$$

Other elements can also be computed in the same way.

The convolution result is

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

b

To keep the size of output feature unchanged, the padding hyper-parameter shall be 1.

The padded feature map is:

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

Using the same method in part a, the result is:

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

(2)

1. For max-pooling the $(0, 0)$ element can be computed by:

$$\max(1, 4, 5, 8) = 8$$

and the result can be similarly computed:

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

2. For average-pooling, the $(0, 0)$ element can be computed by:

$$\frac{1 + 4 + 5 + 8}{4} = 4.5$$

and the result can be similarly computed:

$$\begin{bmatrix} \frac{9}{2} & \frac{5}{2} \\ \frac{17}{4} & 3 \end{bmatrix}$$