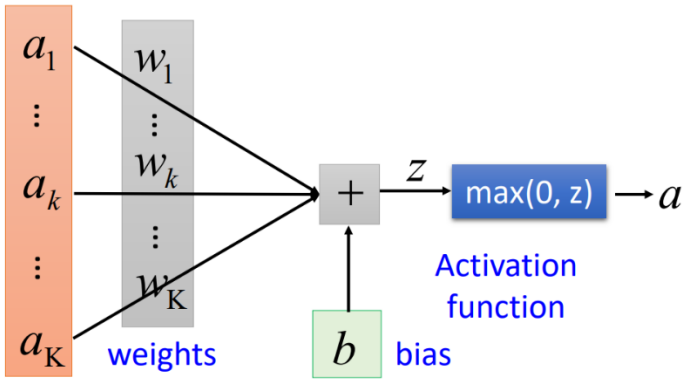
Tutorial 4: Convolutional Neural Networks

1. Explain a) what a neuron is in neural networks, b) how the neurons are designed differently in multilayer perceptions and convolutional neural networks, and c) how neurons are organised to formulate a deep neural network.

   a). A neuron is a linear transformation of the input followed by an activation function (which is usually nonlinear such as ReLu and Sigmoid). The figure illustrates a neuron with ReLU as the activation function. a_i are the features of the input.

$$z = a_1 w_1 + \cdots + a_k w_k + \cdots + a_K w_K + b$$



   b). In MLP, a neuron can be connected to any of the features of the input, while in CNN it is connected to a subwindow of the image and all the neurons in one layer share the same weights (i.e., the parameters in the convolution kernel).

   c) In MLP, multiple neurons connected to the same input formulate a layer of the neural network and multiple layers can be stacked to formulate a deep neural network. In CNN, one convolution formulates a layer and multiple layers are stacked to formulate a deep CNN network. (so far, we consider the case where the input and every layer has one channel. In the next lectures, we will consider the general cases).

2. How many neurons do we have if a 3x3 convolution kernel is applied on a 100x100 image?
   The dimension of the output is 98x98, so we have 98x98 neurons, all of them share the same weights but connected to different subwindows of the image.

3. Suppose a 3x3 convolution kernel is used in the first layer, a 4x4 convolution kernel is used in the second layer and a 5x5 convolution kernel is used in the third layer. What are the receptive fields of these three convolution kernels?

   The receptive field is the number of elements of input image required to access in order to compute the output of the convolution.
   For 3x3 conv in the first layer, it needs to access a 3x3 subwindow of the image so its receptive field is 9.
   For the 4x4 convolution in the second layer, it needs to access a 4x4 subwindow of the output image of the first layer. To compute all these elements, it needs to access a (4+(3-1))x(4+(3-1))=6x6 window of the input image. So its receptive field is 36.
   Similarly, for the 5x5 convolution in the third layer, it need to access a (5+4+3-2))x (5+4+3-2) window of the input image, so its receptive field is 10x10=100.

   Note that, for any nxn image, if a n_1xn_1 kernel is used, the size of the output image will be reduced (n_1-1) rows and (n_1-1) columns hence becomes (n-n_1+1)x(n-n_1+1). For

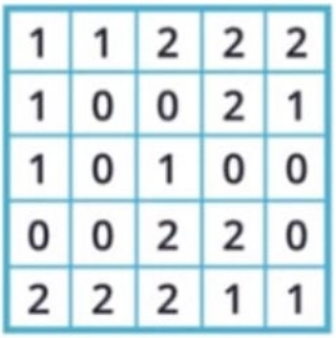
10x10 images, after 3x3 convolution, the output will be of size 8x8. Then after the 4x4 conv in the second layer, the size become 5x5. Finally after the 5x5 conv in the third layer, the size is 1x1=1.

4. Can a linear model be represented with a convolution? If so, what is its receptive field?
   Yes, we can design the kernel with the same size as the input. Hence the receptive field is the dimension of the input.

5. Describe an analogous convolutional layer for audio.

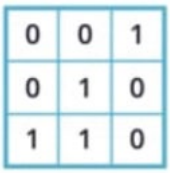
   1xn conv kernel since audio signals are 1 dimentional.

6. Consider the following image



1	1	2	2	2
1	0	0	2	1
1	0	1	0	0
0	0	2	2	0
2	2	2	1	1

   and a convolution kernel



0	0	1
0	1	0
1	1	0

   What is the output of the convolution?

   The top left corner of the output image is

   0x1+0x1+2x1+1x0+0x1+0x0+1x1+0x1+1x0=2+1=3.

   Strictly speaking, we need to split the kernel horizonally and vertically. But as the implementation of convolution uses correlation, we use the correlation here. If questions are raised by students, they can compute the output using either correlation or the strict convolution.