Problem Set 01

Try to solve all questions

Question 01

Neural networks are built out of units with real-valued inputs $X_1 \ldots X_n$, where the unit output Y is given by

$$Y = \frac{1}{1 + e^{-(w_0 + \sum_i w_i X_i)}}$$

Assume that X_i will be 0 or 1. Of course the output Y will be real-valued, ranging anywhere between 0 and 1. We will interpret Y as a boolean value by interpreting it to be a boolean 1 if $Y > 0.5$, and interpreting it to be 0 otherwise.

a) Give 3 weights for a single unit with two inputs X_1 and X_2 , that implements the logical OR function.
b) Can you implement the logical AND function in a single unit? If so, give weights that achieve this. If not, explain the problem.
c) It is impossible to implement the EXCLUSIVE-OR function $Y = X_1 \oplus X_2$ in a single unit. However, you can do it using a multiple unit neural network. Please do. Use the smallest number of units you can. Draw your network, and show all weights of each unit.
d) Create a neural network with only one hidden layer (of any number of units) that implements (*A* OR *B*') $\oplus$ (*C*' OR *D*'). Draw your network, and show all weights of each unit. (*B*' = NOT(*B*)).

Question 02

Why do we need activation functions?

Question 03

Given a Convolution Layer with 8 filters, a filter size of 6, a stride of 2, and a padding of 1. For an input feature map of 32 × 32 × 32, what is the output dimensionality after applying the Convolution Layer to the input?

Question 04

Why are convolutional layers more commonly used than fully-connected layers for image processing?

Question 05

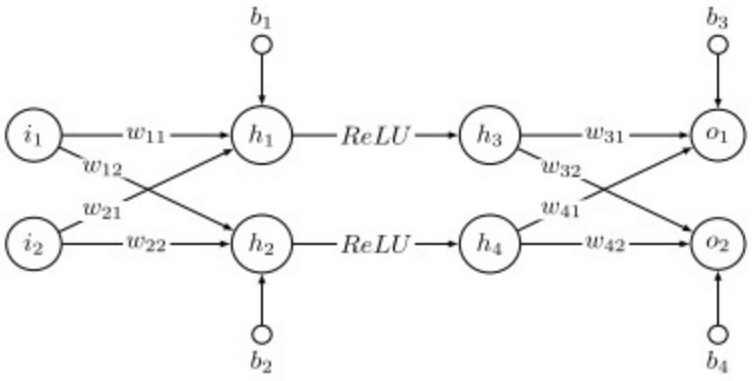
If a data block in a convolutional network has dimension $h \times w \times d = 200 \times 200 \times 128$, and we apply a convolutional filter to it of dimensions $h_f \times w_f \times d_f = 7 \times 7 \times 128$, what is the dimension of the output data block?

Question 06

Name three popular activation functions and draw them?

Question 07

Given the following neural network with fully connection layer and ReLU activations, including two input units (i_1 , i_2), four hidden units (h_1 , h_2) and (h_3 , h_4).The output units are indicated as (o_1 , o_2) and their targets are indicated as (t_1 , t_2). The weights and bias of fully connected layer are called *w* and *b* with specific sub-descriptors.



The values of variables are given in the following table:

Variable	i_1	i_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{31}	w_{32}	w_{41}	w_{42}	b_1	b_2	b_3	b_4	t_1	t_2
Value	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

a) Compute the output (o_1 , o_2) with the input (i_1 , i_2) and network parameters as specified above. Write down all calculations, including intermediate layer results.
b) Compute the mean squared error of the output (o_1 , o_2) calculated above and the target (t_1 , t_2).
c) Update the weight w_{21} using gradient descent with learning rate 0.1 as well as the loss computed previously. (Please write down all your computations.)

Question 08

How many neurons do you need in the output layer if you want to classify email into spam or ham? What activation function should you use in the output layer? If instead you want to tackle MNIST, how many neurons do you need in the output layer, using what activation function? Answer the same questions for getting your network to predict housing prices as discussed in the class.

Question 09

Can you list all the hyperparameters you can tweak in an MLP? If the MLP overfits the training data, how could you tweak these hyperparameters to try to solve the problem?

Question 10

In which cases would you want to use each of the following activation functions: ReLU, tanh, logistic, and softmax?