Quiz 3:
Names: Gabriel Hanson, Shuhao Dong, Nibir Chandra Mandal, Xindi Guo, Uday Kiran
Computing ID: gh8sj, sd7zv, wyr6fx, ptk2ks, rfn3ua

Q1
1.B
2.A
3.A
4.A
5.A
6.B
7.B
8.D
9.A
10.B

Q2
A. 4, 3, 1, 5, 2
B. 1
C. 4
D. 1
E. 4
F. 1
G. 2, 3
H. 1, 2, 3, 4
I. 1, 3, 4
J. 1, 2, 3

Q3

$x_1 = 0$

$x_2 = 1$

$a_1' = 0.5 + (-1) \times 0 + (-1) \times 1$

$= 0.5 - 1 = -0.5$

$a_1 = f(-0.5) = 0$

$a_2' = -1.5 + 1 \times 0 + 1 \times 1$

$= -1.5 + 1 = -05$

$a_2 = f(-0.5) = 0$

$a_3' = -0.5 + 1 \times 0 + 1 \times 0$

$= -0.5$

$a_3 = f(-0.5) = 0$

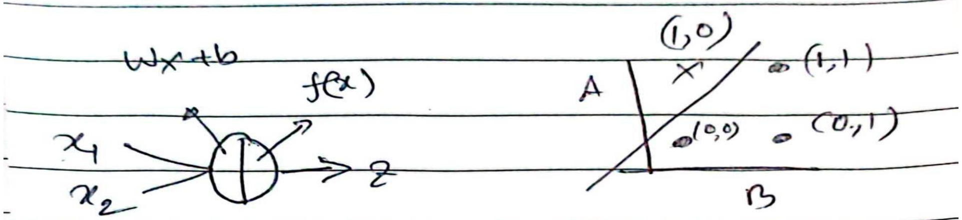
Output of the Network 0

(Ans)

1.

2.

Yes, it can be represented using single unit from neural network.



$$W = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix} \quad b = 0.5$$

$$f(x) = \begin{cases} 1 & \text{when } x = 0 \\ 0 & x \neq 0 \end{cases}$$

when $A=0, B=0,\quad 0+0+0.5=0.5$

$z = f(0.5) = 0$

$A=1, B=1,\quad -0.5+1+0.5 = 1$

$z = f(1) = 0$

$A=0, B=1,\quad 0+1+0.5 = 1.5$

$z = f(1.5) = 0$

$A=1, B=0,\quad -0.5+0+0.5 = 0$

$f(0) = 1$

So, $W = \begin{bmatrix} -0.5 \\ 1 \end{bmatrix}$, $b = 0.5$

Q4

A.

possible reason1: Using dropout regularization in your network could cause the validation error to be smaller than training error.

possible reasons2: Data augmentation only worked on training sets but not on validation sets would cause this error.

B.

Because at high and low values, the derivative of the sigmoid function is very small. When backpropogating these errors, they multiply exponentially and thus values outside the linear region of the sigmoid function either blow up or vanish.

ReLu doesn’t have this problem because the derivatives are either zero or 1 for all values of X. Thus you don’t get the exponential increase or decrease that sigmoid creates.

C.
Activation functions play an integral role in neural networks by introducing nonlinearity. This nonlinearity allows neural networks to develop complex representations and functions based on the inputs that would not be possible with a simple linear regression model.

D.
In image processing we are interested in finding patterns in the images. In fully connected networks each pixel is connected to each other pixel. This is not necessary in image processing as patterns extend to adjacent pixels but not jumping across pixels to far away ones. Thus a convolutional filter allows for the recognition of patterns in subsets of adjacent pixels.
Also fully connected networks will have a higher number of weights to train that implies more training time.

E.
For many image processing problems, the patterns within the images themselves are similar, but which patterns that are important change. The initial layers in these networks are complex and computationally expensive as they learn the patterns within the images, but these patterns themselves are usually constant for another task. When we fine tune for another task we are only adjusting some of the weights to decipher the importance of the patterns for our new problem.

F.
I prefer to train a single neural network with individual outputs and a shared hidden layer. It will be less computationally expensive and our individual output neuron will extract the features necessary for classification of that specific disease. The final layer of neurons will not communicate with each other anyway.