1. When $tanh(x)=T$ and $sigmoid(x)=S$ show that: [2]

$$T = \frac{2S-1}{2S^2-2s+1}$$

$$S = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$$

$$S + Se^x = e^x \Rightarrow e^x = \frac{S}{1-S}$$

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} = \frac{(e^x)^2-1}{(e^x)^2+1}$$

Substituting for e^x

$$\Rightarrow \frac{\left(\frac{S}{1-S}\right)^2-1}{\left(\frac{S}{1-S}\right)^2+1}$$

$$\Rightarrow \frac{S^2-(1-S)^2}{S^2+(1-S)^2}$$

$$\therefore\ \tanh(x) = \frac{2S-1}{2S^2-2S+1}$$

2. A deep neural network for a 4-class classification problem has, for a particular input, the pre-activation vector at output layer as [-1 0 5 3] corresponding to classes [A B C D]. Which class will the model predict for the input? [2]

z = [-1 0 5 3]

e^z = [0.368 1 148.41 20.09]

sum = 169.868

Softmax(z) = $e^z/\text{sum}(e^z)$ = [0.002 0.006 0.873 0.118]

Model will predict Class C.

3. Consider the following neural network for binary classification [0 1]. The input is $[0.1 \quad 0.5]^T$ and belongs to Class 1. Use an appropriate loss function.

a) Compute the loss at the output node assuming hidden layer uses sigmoid activation function. [2]

$z_1^1 = 0.1x_1 + 0.3x_2 + 0.25 = 0.41$

$a_1^1 = \text{sigmoid}(z_1^1) = 0.601$

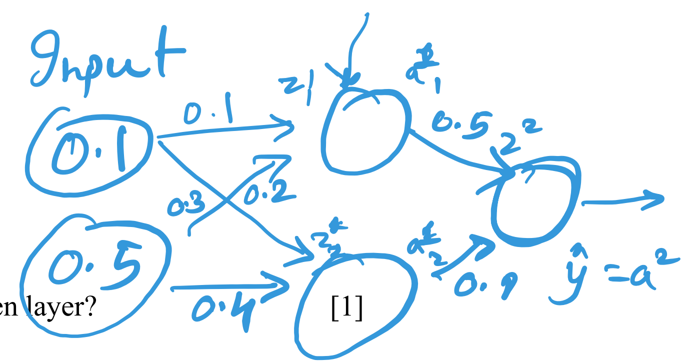
$z_2^1 = 0.2x_1 + 0.4x_2 + 0.25 = 0.47$

$a_2^1 = \text{sigmoid}(z_2^1) = 0.615$

$z^2 = 0.5\, a_1^1 + 0.8\, a_2^1 + 0.35 = 1.1425$

y' = $\text{sigmoid}(z^2)$ = 0.758

Loss = -log y' = 0.277



b) What will the loss if ReLU is used in the hidden layer? [1]

$z_1^1 = 0.41$

$a_1^1 = \text{ReLU}(z_1^1) = 0.41$

$z_2^1 = 0.47$

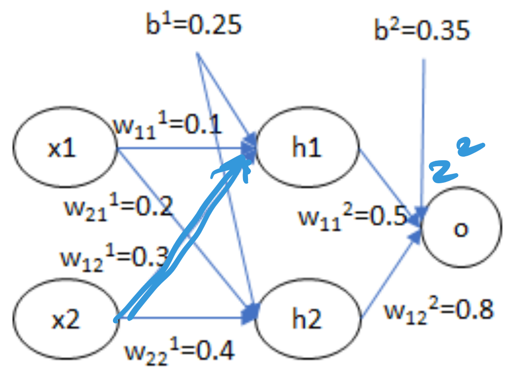
$a_2^1 = \text{ReLU}(z_2^1) = 0.47$

$z^2 = 0.5\, a_1^1 + 0.8\, a_2^1 + 0.35 = 0.931$

y' = $\text{sigmoid}(z^2)$ = 0.717

Loss = -log y' = 0.3327

c) Derive the updated weight w_{12}^1 after one iteration. w_{ij}^k refers to weight of connection between ith neuron in layer k with jth neuron of layer k-1. Assuming ReLU activation in hiddel layer and a learning rate of α= 0.1. [3]



$$\frac{\partial L}{\partial w_{12}^1} = \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial z^2} * \frac{\partial z^2}{\partial a_1^1} * \frac{\partial a_1^1}{\partial z_1^1} * \frac{\partial z_1^1}{\partial w_{12}^1}$$

$$= \frac{-1}{\hat{y}} * \hat{y}(1-\hat{y}) * w_{11}^2 * 1 * x_2$$

$$= -(1-\hat{y}) * w_{11}^2 * x_2$$

$$w_{11}^2(new) = w_{11}^2(old) - 0.1 * \frac{\partial L}{\partial w_{12}^1}$$

$$w_{11}^2(new) = 0.3 - 0.1 * [-(1-0.7402) * 0.5 * 0.5]$$
$$w_{11}^2(new) = 0.306$$

4. Give brief answers to the following questions:
   a) A deep neural network has 100 hidden layers. How can the depth affect the learning and performance of the network?
   b) How does dropout help in increasing performance of deep neural networks?
   c) "Using L1 loss enforces sparsity on the weights of the network." Do you agree with this statement? Why/Why not?
   d) You train a deep neural network with a two hidden layers and observe that training and validation accuracy is low. You increase the number of hidden layers. Will this solve the issue or will there be a problem? Explain.
   e) Consider the Loss vs. Iterations plot given below. Will early stopping technique be useful in this case? Justify.