This quiz covers units 4 and 5. Each question is worth 25%. Please make sure to follow the hand-in instructions described in Canvas announcements and in the course website.

Question 1: Consider a small fully connected network with input vectors of size 10. Assume 8 neurons in the first hidden layer with Sigmoid activation and assume 2 neurons in the second hidden layer with ReLu activation. Finally assume 5 neurons in the output layer, operated on by Softmax.

(a) Determine the number of parameters in this network.

(b) Assume now that after training, you observe that each and every one of the training and validation images drives the inputs to the ReLu activations of the second layer to be positive (and hence these ReLu activations act as identity functions). Assume the trained weight matrices of the layers are $W^{[1]}$, $W^{[2]}$, and $W^{[3]}$ and the bias vectors are $b^{[1]}$, $b^{[2]}$, and $b^{[3]}$, suggest an equivalent network architecture with only two layers and represent the parameters of this network in terms of $W^{[i]}$ and $b^{[i]}$.

Solution:

(a) This is simply a count of the connections (weights) and biases per layer:

$$(8 \times 10 + 8) + (2 \times 8 + 2) + (5 \times 2 + 5) = 121.$$

(b) The alternative network has a first hidden layer with weight parameters $W^{[1]}$ and bias parameters $b^{[1]}$ just like the original network. However, unlike the original network that has three layers, the alternative network only has two layers. In the original network, if the output of the first layer is the 8-vector u, the input to the Softmax is,

$$W^{[3]}(W^{[2]}u + b^{[2]}) + b^{[3]} = \underbrace{W^{[3]}W^{[2]}}_{\widetilde{W}^{[2]}} u + \underbrace{W^{[3]}b^{[2]} + b^{[3]}}_{\widetilde{b}^{[2]}}.$$

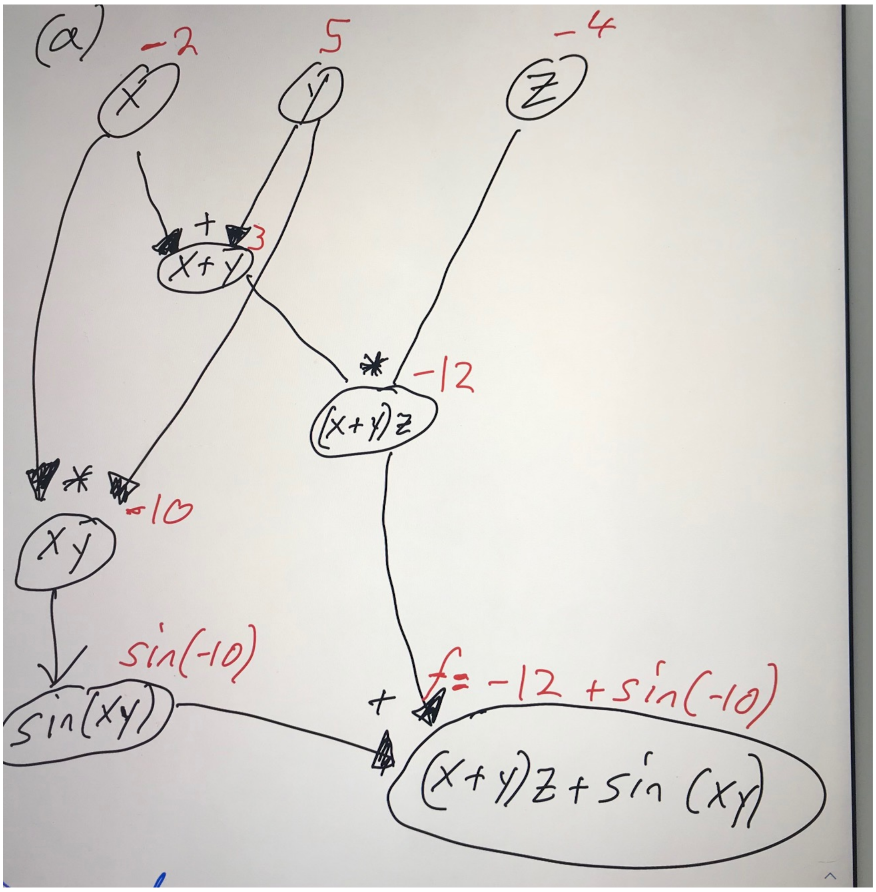
Hence the second (and final) layer of the alternative network has a 5×8 weight matrix $\widetilde{W}^{[2]}$ as above, and bias vector of length 5, which is $\widetilde{b}^{[2]}$ as above.

Question 2:
Consider the following function $f(x, y, z) = (x + y)z + \sin(xy)$.

(a) Draw the computational graph associated to this function by decomposing the function into elementary operations.

(b) Let consider the entries $x = -2$, $y = 5$, $z = -4$. Perform a forward pass and a backward pass in order to get $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$. You may leave your answer in non-numeric form (e.g. no need to numerically evaluate trigonometric functions). Also remember that $\sin'(z) = \cos(z)$.

Solution:



(b) The forward pass is already presented in red in the illustration above. The backward pass operates as follows:

$$
\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{\partial f}{\partial (x+y)z} \cdot \frac{\partial (x+y)z}{\partial (x+y)} \cdot \frac{\partial (x+y)}{\partial x} + \frac{\partial f}{\partial \sin(xy)} \cdot \frac{\partial \sin(xy)}{\partial xy} \cdot \frac{\partial xy}{\partial x} \\
&= 1 \cdot z \cdot 1 \qquad + \qquad 1 \cdot \cos(xy) \cdot y \\
&= 1 \cdot (-4) \cdot 1 \qquad + \qquad 1 \cdot \cos(-10) \cdot 5 \\
&= 5\cos(-10) - 4
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial f}{\partial y} &= \frac{\partial f}{\partial (x+y)z} \cdot \frac{\partial (x+y)z}{\partial (x+y)} \cdot \frac{\partial (x+y)}{\partial y} + \frac{\partial f}{\partial \sin(xy)} \cdot \frac{\partial \sin(xy)}{\partial xy} \cdot \frac{\partial xy}{\partial y} \\
&= 1 \cdot z \cdot 1 \qquad\qquad\qquad\qquad\qquad\qquad + \qquad\qquad 1 \cdot \cos(xy) \cdot x \\
&= 1 \cdot (-4) \cdot 1 \qquad + \qquad 1 \cdot \cos(-10) \cdot (-2) \\
&= -2\cos(-10) - 4
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f}{\partial z} &= \frac{\partial f}{\partial (x+y)z} \cdot \frac{\partial (x+y)z}{\partial z} + \frac{\partial f}{\partial \sin(xy)} \cdot \frac{\partial \sin(xy)}{\partial z} \\
&= 1 \cdot (x+y) \qquad\qquad\qquad + \qquad\qquad 1 \cdot \cos(xy) \cdot 0 \\
&= 1 \cdot 3 \qquad + \qquad 0 \\
&= 3
\end{aligned}$$

Question 3: Consider a convolutional neural network operating on input images of size 100×100 and depth 3 (RGB).

1. The first layer has 5×5 convolutions with padding $= 2$ (in each direction) a stride of 1, and 8 output channels (one filter per channel). The layer also has 2×2 max-pooling and a ReLu activation function.
2. The second layer has 3×3 convolutions with stride $= 1$, padding $= 1$, 16 output channels, no max-pooling, and ReLu activation.
3. The third layer has 1×1 convolutions with stride $= 1$, padding $= 0$, 4 output channels, 2×2 max-pooling, and ReLu activation.
4. The fourth layer is a fully connected layer with 20 neurons with a weight matrix $W^{[4]}$ of dimension $20 \times N$, and ReLu activation.
5. The fifth layer is a fully connected layer with 3 neurons, and is followed by a softmax.

(a) Determine the value of N.

(b) Determine the number of trained parameters in the network.

(c) Determine the number of neurons in the network (every pixel in a convolutional channel counts as a neuron).

Solution:

(a) Output dimension of layer 1 (after max-pooling):

$$\frac{1}{2}\Big(\frac{100 \ + \ 2\times 2 \ - \ 5}{1} + 1\Big) = 50.$$

Output dimension of layer 2 :

$$\frac{50 \ + \ 2\times 1 \ - \ 3}{1} + 1 = 50.$$

Output dimension of layer 3 is 25 because it has 1×1 convolutions and 2×2 max-pooling. With 4 channels this is $25 \times 25 \times 4 = 2500$ neurones.

Hence the weight matrix $W^{[4]}$ is of dimension 20×2500.

(b) The number of trained parameters is $51,927$. Here is the calculation:

$$\underbrace{8 \times (5 \times 5 \times 3 \ + \ 1)}_{\text{layer 1}} + \underbrace{16 \times (3 \times 3 \times 8 \ + \ 1)}_{\text{layer 2}} + \underbrace{4 \times (1 \times 1 \times 16 \ + \ 1)}_{\text{layer 3}} + \underbrace{(20 \times 2500 + 20)}_{\text{layer 4}} + \underbrace{(3 \times 20 + 3)}_{\text{layer 5}}.$$

(c) The number of neurons is $62,523$. Here is the calculation:

$$\underbrace{8 \times 50 \times 50}_{\text{layer 1}} + \underbrace{16 \times 50 \times 50}_{\text{layer 2}} + \underbrace{4 \times 25 \times 25}_{\text{layer 3}} + \underbrace{20}_{\text{layer 4}} + \underbrace{3}_{\text{layer 5}}.$$

Question 4: Consider a fully connected (dense) network with input vectors of size 10. The network has 50 neurons in the first hidden layer and 5 neurons in the second layer which is the output layer. Say you now decide to treat it as a convolutional network using 1×1 convolutions. For this you represent the input as a $1 \times 1 \times 10$ tensor.

Describe the layers of the network, making sure each layer is a convolutional layer.

Solution:

The first layer is a $1 \times 1 \times 10$ convolutional layer with 50 channels (50 such filters). This creates a volume of $1 \times 1 \times 50$.

The second layer is a $1 \times 1 \times 50$ convolutional layer with 5 output channels (5 such filters). This creates a volume of $1 \times 1 \times 5$.