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 $b^{[i]}$.

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) 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 $\sin'(z) = \cos(z)$.

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).

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.