CS6423: Scalable Computing **Name:** ______________________

Spring 2020

Mid Term Practice Questions

Time Limit: Student ID ______________________

---

1. (20 points) You design the 2-layer fully connected neural network shown in Figure 1. All activations are sigmoids and your optimizer is stochastic gradient descent. The loss function is $(\hat{y} - y)^2$. in the network.

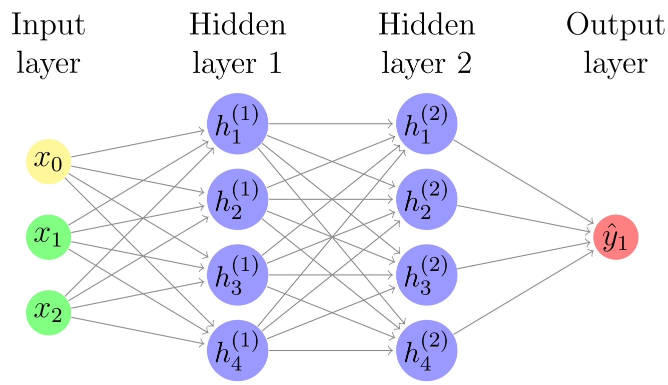

Figure 1: Simple deep network with sigmoid activations

   (a) (10 points) Draw the computation graph for this example.

   (b) (5 points) Compare how Caffe and TensorFlow would implement this example.

   (c) (5 points) Compare the assignment of CPU/GPU processors for Caffe and TensorFlow.

2. (20 points) We have a deep network with input x, target label t output y, hidden nodes z, and loss $\mathcal{L}$ defined by the following:

$$\begin{aligned} h &= W_1 x + b \\ \hat{y} &= ReLU(W_2 h) \\ \mathcal{L} &= 1/2(y - t)^2 \end{aligned}$$

where ReLU(x) is $\max(0, x)$ for $x > 0$ and 0 otherwise. See the neural network shown in Figure 2

   (a) (10 points) Draw the computation graph for this example.

   (b) (5 points) If we initialise $W_1 = 0$, $b = [1\ 0\ 0\ 0]^T$, $W_2 = [1\ 1\ 1_1]^T$ and input $x = [1\ 1\ 0]^T$ with target $t = 2$, compute the loss for this network.

   (c) (5 points) Compute how the weights W, b are updated for a learning rate $\eta = 1$ for this first iteration of backpropagation.

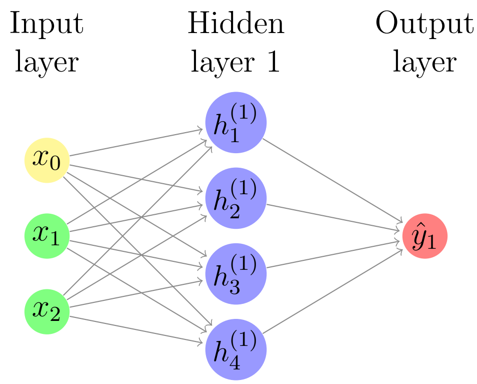

Figure 2: Simple deep network with sigmoid activations

Solution:

$$
\begin{aligned}
h &= W_1 x + b = [0][1\ 1\ 1]^T + [1\ 0\ 0\ 0]^T \ = \ [1\ 0\ 0\ 0]^T \\
\hat{y} &= ReLU(W_2 h) \ = \ ReLU\left[[1\ 1\ 1\ 1]^T \cdot [1\ 0\ 0\ 0]\right] = 1 \\
\mathcal{L} &= 1/2(\hat{y} - t)^2 \ = \ 1/2(1-2)^2 = 1/2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial y} &= \frac{\partial}{\partial y}\left[1/2(y-t)^2\right] = (y-t) = (1-2) = -1 \\
\frac{\partial L}{\partial h} &= \frac{\partial L}{\partial y}\frac{\partial y}{\partial h} \ = \ (y-t)[1\ 1\ 1\ 1]^T \text{ if } h > 0 = -1 * [1\ 1\ 1\ 1]^T = [-1\ \ -1\ \ -1\ \ -1]^T \\
\frac{\partial L}{\partial W} &= \frac{\partial L}{\partial h}\frac{\partial h}{\partial W} \\
\frac{\partial L}{\partial b} &= \frac{\partial L}{\partial h}\frac{\partial h}{\partial b} \ = \ \frac{\partial L}{\partial h}[1\ 1\ 1\ 1] = [-1\ \ -1\ \ -1\ \ -1]
\end{aligned}
$$