## CS 7150, in-class quiz 1 Your Name:

1. What do you think of the pace of the course so far? (5 points)
   (a) Too slow. (b) Just right. (c) Too fast.

2. Which of the following would you consider to be valid activation functions (performed in an element-wise manner) to train neural networks. (12 points)

   (a) $f(x) = \max(0, x)$

   (b) $f(x) = x + 1$

   (c) $f(x) = \begin{cases} \min(x, 0.1x), & \text{if } x \geq 0 \\ \min(x, 0.1x), & \text{if } x < 0 \end{cases}$

   (d) $f(x) = \begin{cases} \max(x, 0.1x), & \text{if } x \geq 0 \\ \min(x, 0.1x), & \text{if } x < 0 \end{cases}$

3. What can you do if you see **underfitting**? (12 points)

   (a) Increase the amount of training data

   (b) Increase the number of model parameters

   (c) Use a large learning rate

   (d) Reduce the number of epochs to train the model

4. What can you do if you see **overfitting**? (12 points)

   (a) Increase the amount of training data

   (b) Increase the number of model parameters

   (c) Use a large learning rate

   (d) Reduce the number of epochs to train the model

5. Alice is designing a fully-connected neural network. But somehow she forgets to add non-linear activation functions in-between. Why may happen to the neural network? (12 points)

6. Name two advantages of using the SGD optimizer with momentum over without momentum. (12 points)

7. You come across the following activation function,

$$f(x) = \begin{cases} 1, \text{ if } x \geq 0, \\ 0, \text{ if } x < 0. \end{cases} \tag{1}$$

A friend highly recommends you to use it in your neural network together with the Stochastic Gradient Descent (SGD) optimizer plus momentum. Would you follow their advise? Why or why not? (15 points)

8. Suppose we are developing a model for binary classification with the cross-entropy loss. For a particular training sample, the ground-truth label is 0. Denote the loss value as $L$ and logits as $s_0$ and $s_1$, respectively. Derive $\frac{\partial L}{\partial s_0}$ and $\frac{\partial L}{\partial s_1}$. (20 points)