NAME: _____

STUDENT ID: _____

Directions:
- ❏ This is an open-book, open-note and open-internet exam.
- ❏ You must answer the questions in your own words.
- ❏ Your answers should be brief, concise and readable.
- ❏ You have 120 minutes to complete the exam. Good luck!

1. (3 pts) Name **THREE** major factors that contribute to the recent success of deep learning.

2. (3 pts) True or False:

    a. _____ Deep learning works well even when we have a small training set.

    b. _____ Training a deep network requires a lot of computational power.

    c. _____ A deep network is basically a neural network with many layers.

3. (3 pts) Consider a neural network that outputs $\hat{y} = \{0,1\}$ and was trained on a dataset $\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\}$. Can you suggest **a loss function** and **a cost function** that we can use for this network? You must write out the functions mathematically using notations provided in the context.

4. (3 pts) What is vectorization and why is it important in deep learning?

5. (3 pts) Write down **THREE** activation functions that are commonly used in deep networks, along with their derivatives.

6. (3 pts) Consider the following function.
$$f(x) = \begin{cases} 0 \text{ if } x < 0 \\ 1 \text{ otherwise} \end{cases}$$

Can we use this function as an activation function? If not, explain why not

7. (3 pts) Why do we need non-linear activation functions in a neural network?

8. (3 pts) What is the learning rate? Where do we use it?

9. (3 pts) In logistic regression, we can initialize all the weights with zeros but, in neural network, we cannot do so. Why?

10. (3 pts) Explain why it is not recommended to initialize weights of a neural network with large values (large positive or small negative).

11. (3 pts) Explain the difference between train set, dev set, and test set.

12. (6 pts) Consider the following situations. Indicate whether it is overfitting, underfitting or something else. Also, suggest what to do next to improve performance.

   a. Bayes error: 2%, Training error: 10%, Test error: 12%

       Choose one or write your answer: (overfitting, underfitting, _____)
       Your suggestion:

   b. Bayes error: 5%, Training error: 1%, Test error 15%

       Choose one or write your answer: (overfitting, underfitting, _____)
       Your suggestion:

   c. Bayes error: 7%, Training error: 6%, Dev error: 7%, Test error: 20%

       Choose one or write your answer: (overfitting, underfitting, _____)
       Your suggestion:

13. (3 pts) Explain what is vanishing/exploding gradient problem? How can we avoid it?

14. (3 pts) What is dropout? How does it help regularizing a network?

15. (4 pts) How many weight updates the following algorithms perform on ONE pass over the same training data? Assume 100 training examples.

    a. _____ Stochastic Gradient Descent
    b. _____ Batch Gradient Descent
    c. _____ Mini-Batch Gradient Descent (batch size = 10)
    d. _____ Mini-Batch Gradient Descent (batch size = 50)

16. (3 pts) Explain how learning rate decay could help in the training process.

17. (3 pts) What is Batch Normalization? What are the benefits of batch normalization?