

✔ Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 80% or higher

Go to next item

1. If you have 20,000,000 examples, how would you split the train/dev/test set? Choose the best option.

1 / 1 point

↗ **Expand**

✔ **Correct**

Yes. Given the size of the dataset, 0.5% of the samples are enough to get a good estimate of how well the model is doing.

2. The dev and test set should:

1 / 1 point

 **Expand**

 **Correct**

3. If your Neural Network model seems to have high bias, what of the following would be promising things to try? (Check all that apply.)

1 / 1 point

 **Expand** **Correct**

Great, you got all the right answers.

- 4.** You are working on an automated check-out kiosk for a supermarket, and are building a classifier for apples, bananas and oranges. Suppose your classifier obtains a training set error of 0.5%, and a dev set error of 7%. Which of the following are promising things to try to improve your classifier? (Check all that apply.)

1 / 1 point **Expand** **Correct**

Great, you got all the right answers.

5. Which of the following are regularization techniques?

1 / 1 point

 **Expand**

 **Correct**

Great, you got all the right answers.

6. The regularization hyperparameter must be set to zero during testing to avoid getting random results. True/False?

1 / 1 point

 **Expand****Correct**

Correct. The regularization parameter affects how the weights change during training, this means during backpropagation. It has no effect during the forward propagation that is when predictions for the test are made.

7. With the inverted dropout technique, at test time:

1 / 1 point **Expand****Correct**

8. During training a deep neural network that uses the tanh activation function, the value of the gradients is practically zero. Which of the following is most likely to help the vanishing gradient problem?

1 / 1 point **Expand****Correct**

Correct. A careful initialization can help reduce the vanishing gradient problem.

9. Which of these techniques are useful for reducing variance (reducing overfitting)?
(Check all that apply.)

1 / 1 point

 **Expand**

 **Correct**

Great, you got all the right answers.

10. Suppose that a model uses, as one feature, the total number of kilometers walked by a person during a year, and another feature is the height of the person in meters. What is the most likely effect of normalization of the input data?

1 / 1 point

 **Expand**

 **Correct**

Correct. Since the difference between the ranges of the features is very different, this will likely cause the process of gradient descent to oscillate, making the optimization process longer.