Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

1. If you have 10,000 examples, how would you split the train/dev/test set? Choose the best option. **0 / 1 point**

- ○ 33% train. 33% dev. 33% test.
- ○ 98% train. 1% dev. 1% test.
- ◉ 60% train. 20% dev. 20% test.

Expand

Incorrect
You answered when the time elapsed.

2. In a personal experiment, an M.L. student decides to not use a test set, only train-dev sets. In this case which of the following is true? **1 / 1 point**

- ◉ He might be overfitting to the dev set.
- ○ He won't be able to measure the bias of the model.
- ○ Not having a test set is unacceptable under any circumstance.
- ○ He won't be able to measure the variance of the model.

Expand

Correct
Yes. Although not recommended, if a more accurate measure of the performance is not necessary it is ok to not use a test set. However, this might cause an overfit to the dev set.

3. If your Neural Network model seems to have high variance, what of the following would be promising things to try? **1 / 1 point**

- ☐ Increase the number of units in each hidden layer
- ☑ Add regularization

  ✓ Correct
- ☐ Get more test data
- ☑ Get more training data

  ✓ Correct
- ☐ Make the Neural Network deeper

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. **0 / 1 point**

bananas, and oranges. Suppose your classifier obtains a training set error of 19% and a dev set error of 21%. Which of the following are promising things to try to improve your classifier? (Check all that apply, suppose the human error is approximately 0%)

- ○ Use a bigger network.
- ◉ Get more training data.
- ○ Increase the regularization parameter lambda.

Expand

✗ **Incorrect**
No. This won't help to reduce the high bias of the model; it is better to address that first before moving to reduce a high variance.

5. Which of the following are regularization techniques? **1 / 1 point**

- ☑ Dropout.

✓ **Correct**
Correct. Using dropout layers is a regularization technique.

- ☐ Gradient Checking.
- ☑ Weight decay.

✓ **Correct**
Correct. Weight decay is a form of regularization.

- ☐ Increase the number of layers of the network.

Expand

✓ **Correct**
Great, you got all the right answers.

6. To reduce high variance, the regularization hyperparameter lambda must be increased. True/False? **1 / 1 point**

- ◉ True
- ○ False

Expand

✓ **Correct**
Correct. By increasing the regularization parameter the magnitude of the weight parameters is reduced. This helps reduce the variance.

7. Which of the following are true about dropout? **1 / 1 point**

- ☑ It helps to reduce the variance of a model.

✓ **Correct**
Correct. The dropout is a regularization technique and thus helps to reduce the variance.

- ☑ In practice, it eliminates units of each layer with a probability of 1- keep_prob.

✓ **Correct**
Correct. The dropout is a regularization technique and thus helps to reduce the overfit.

- ☐ In practice, it eliminates units of each layer with a probability of keep_prob.

- [ ] It helps to reduce the bias of a model.

Expand

Correct
Great, you got all the right answers.

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**

- ( ) Increase the number of layers of the network.
- (•) Use Xavier initialization.
- ( ) Use a larger regularization parameter.
- ( ) Increase the number of cycles during the training.

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**

- [ ] Xavier initialization
- [ ] Exploding gradient
- [x] Dropout

Correct

- [ ] Vanishing gradient
- [x] Data augmentation

Correct

- [x] L2 regularization

Correct

- [ ] Gradient Checking

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**

- ( ) It will make the data easier to visualize.
- ( ) It will increase the variance of the model.
- (•) It will make the training faster.
- ( ) It won't have any positive or negative effects.

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.