Your grade: 100%

Your latest: **100%** • Your highest: **100%**
To pass you need at least 80%. We keep your highest score.

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**

- (●) 99% train. 0.5% dev. 0.5% test.
- ( ) 90% train. 5% dev. 5% test.
- ( ) 60% train. 20% dev. 20% test.

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**

- ( ) Come from different distributions
- (●) Come from the same distribution
- ( ) Have the same number of examples
- ( ) Be identical to each other (same (x,y) pairs)

Correct

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

- [x] Get more training data

Correct

- [ ] Get more test data
- [x] Add regularization

Correct

- [ ] Make the Neural Network deeper
- [ ] Increase the number of units in each hidden layer

4. You are working on an automated check-out kiosk for a supermarket and are building a classifier for apples, bananas, and oranges. **1 / 1 point**

Suppose your classifier obtains a training set error of 19% and a development set error of 21%.

Which of the following is the most promising strategy to improve your classifier? (Assume the human error is approximately 0%)

- (•) Use a bigger network.
- ( ) Increase the regularization parameter lambda.
- ( ) Get more training data.

Correct
A larger network can reduce bias by enabling the model to learn more complex patterns.

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

- [x] Dropout.

Correct
Correct. Using dropout layers is a regularization technique.

- [ ] Gradient Checking.
- [ ] Increase the number of layers of the network.
- [x] Weight decay.

Correct
Correct. Weight decay is a form of regularization.

6. **True or False:** In L2 regularization, the lambda hyperparameter directly influences the calculations used by the model to make predictions during testing. **1 / 1 point**

- ○ True
- ◉ False

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**

- ○ You apply dropout (randomly eliminating units) but keep the 1/keep_prob factor in the calculations used in training.

- ○ You apply dropout (randomly eliminating units) and do not keep the 1/keep_prob factor in the calculations used in training
- ◉ You do not apply dropout (do not randomly eliminate units) and do not keep the 1/keep_prob factor in the calculations used in training
- ○ You do not apply dropout (do not randomly eliminate units), but keep the 1/keep_prob factor in the calculations used in training.

✓ **Correct**

8. Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply) **1 / 1 point**

- ☐ Increasing the regularization effect
- ☑ Reducing the regularization effect

✓ **Correct**

- ☐ Causing the neural network to end up with a higher training set error
- ☑ Causing the neural network to end up with a lower training set error

✓ **Correct**

9. Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) **1 / 1 point**

- ☐ Exploding gradient
- ☐ Xavier initialization
- ☐ Vanishing gradient

- [x] Data augmentation

Correct

- [x] Dropout

Correct

- [ ] Gradient Checking
- [x] L2 regularization

Correct

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?

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

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.