

## 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 10,000 examples, how would you split the train/dev/test set? Choose the best option.

1 / 1 point

- ☒ 60% train. 20% dev. 20% test.
- ☐ 98% train. 1% dev. 1% test.
- ☐ 33% train. 33% dev. 33% test.

✓ Correct

Yes. This might be considered a small data set, not in the range of big data. Thus a more classical (old) best practice should be used.

2. When designing a neural network to detect if a house cat is present in the picture, 500,000 pictures of cats were taken by their owners. **These are used to make the training, dev and test sets.** It is decided that to increase the size of the test set, 10,000 new images of cats taken from security cameras are going to be used in the test set. Which of the following is true?

- ☐ This will reduce the bias of the model and help improve it.
- ☐ This will increase the bias of the model so the new images shouldn't be used.
- ☒ This will be harmful to the project since now dev and test sets have different distributions.

✓ Correct

Yes. The quality and type of images are quite different thus we can't consider that the dev and the test sets came from the same distribution.

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

☒ Add regularization

✓ Correct

☐ Increase the number of units in each hidden layer

☒ Get more training data

✓ Correct

☐ Get more test data

☐ Make the Neural Network deeper

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 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%)

- ☐ Increase the regularization parameter lambda.
- ☒ Use a bigger network.
- ☐ Get more training data.

✓ Correct

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

5. What is weight decay?

- ☐ Gradual corruption of the weights in the neural network if it is trained on noisy data.
- ☐ The process of gradually decreasing the learning rate during training.
- ☐ A technique to avoid vanishing gradient by imposing a ceiling on the values of the weights.
- ☒ A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration.

✓ Correct

6. What happens when you increase the regularization hyperparameter lambda?

- ☒ Weights are pushed toward becoming smaller (closer to 0)
- ☐ Gradient descent taking bigger steps with each iteration (proportional to lambda)
- ☐ Doubling lambda should roughly result in doubling the weights
- ☐ Weights are pushed toward becoming bigger (further from 0)

✔ Correct

7. Which of the following are true about dropout?

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

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

- ☐ It helps to reduce the bias of a model.
- ☐ In practice, it eliminates units of each layer with a probability of keep\_prob.
- ☒ 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.

8. Increasing the parameter keep\_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply)

- ☐ 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 the following actions increase the regularization of a model? (Check all that apply)

- ☒ Make use of data augmentation.

✔ Correct  
Correct. Data augmentation has a way to generate "new" data at a relatively low cost. Thus making use of data augmentation can reduce the variance.

- ☐ Increase the value of keep\_prob in dropout.
- ☒ Increase the value of the hyperparameter lambda.

✔ Correct  
Correct. When increasing the hyperparameter lambda we increase the effect of the L<sub>2</sub> penalization.

- ☐ Normalizing the data.
- ☐ Decrease the value of the hyperparameter lambda.

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 won't have any positive or negative effects.
- ☒ It will make the training faster.
- ☐ It will increase the variance of the model.

✔ 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.