

Your grade: 90%

Your latest: 90% • Your highest: 90% • 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

- ☐ 98% train. 1% dev. 1% test.
- ☒ 60% train. 20% dev. 20% 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. 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 won't be able to measure the variance of the model.
- ☐ Not having a test set is unacceptable under any circumstance.
- ☐ He won't be able to measure the bias of the model.
- ☒ He might be overfitting to the dev set.

✓ 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

- ☐ Get more test data
- ☒ Add regularization

✓ Correct

- ☐ Make the Neural Network deeper
- ☐ Increase the number of units in each hidden layer
- ☒ Get more training data

✓ Correct

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

☒ Increase the regularization parameter λ

✓ Correct

☐ Decrease the regularization parameter λ

☒ Get more training data

✓ Correct

☐ Use a bigger neural network

5. In every case it is a good practice to use dropout when training a deep neural network because it can help to prevent overfitting. True/False?

1 / 1 point

☐ True

☒ False

✓ Correct

Correct. In most cases, it is recommended to not use dropout if there is no overfit. Although in computer vision, due to the nature of the data, it is the default practice.

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

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:

0 / 1 point

- ☒ You do not apply dropout (do not randomly eliminate units), but keep the $1/\text{keep_prob}$ factor in the calculations used in training.
- ☐ You apply dropout (randomly eliminating units) and do not keep the $1/\text{keep_prob}$ factor in the calculations used in training
- ☐ You do not apply dropout (do not randomly eliminate units) and do not keep the $1/\text{keep_prob}$ factor in the calculations used in training
- ☐ You apply dropout (randomly eliminating units) but keep the $1/\text{keep_prob}$ factor in the calculations used in training.

✗ Incorrect

8. Decreasing the parameter `keep_prob` from (say) 0.6 to 0.4 will likely cause the following:

1 / 1 point

- ☐ Causing the neural network to have a higher variance.
- ☒ Increasing the regularization effect.
- ☐ Reducing the regularization effect.

✓ Correct

Correct. This will make the dropout have a higher probability of eliminating a node in the neural network, increasing the regularization effect.

9. Which of the following actions increase the regularization of a model? (Check all that apply)

1 / 1 point

- ☐ Decrease the value of the hyperparameter `lambda`.
- ☐ Normalizing the data.
- ☒ 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_2 penalization.

1 / 1 point

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.