

Your grade: 100%

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

- ☐ 33% train. 33% dev. 33% test.
- ☐ 98% train. 1% dev. 1% test.
- ☒ 60% train. 20% dev. 20% 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.
- ☐ He won't be able to measure the bias of the model.
- ☒ He might be overfitting to the dev set.
- ☐ Not having a test set is unacceptable under any circumstance.

✔ 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

- ☒ Add regularization

✔ Correct

- ☐ Get more test data
- ☒ Get more training data

✔ Correct

- ☐ Increase the number of units in each hidden layer
- ☐ 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 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 lambda

✔ Correct

- ☐ Decrease the regularization parameter lambda
- ☒ Get more training data

✔ Correct

- ☐ Use a bigger neural network

5. What is weight decay?

1 / 1 point

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

✔ Correct

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

1 / 1 point

- ☐ Weights are pushed toward becoming bigger (further from 0)
- ☒ 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

✔ Correct

7. Which of the following are true about dropout?

1 / 1 point

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

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

✔ Correct

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

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 a larger regularization parameter.
- ☒ Use Xavier initialization.
- ☐ Increase the number of cycles during the training.

✔ 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

- ☒ Data augmentation

✔ Correct

- ☒ Dropout

✔ Correct

- ☒ L2 regularization

✔ Correct

- ☐ Gradient Checking
- ☐ Xavier initialization
- ☐ Exploding gradient
- ☐ Vanishing gradient

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 won't have any positive or negative effects.
- ☒ It will make the training faster.

✔ 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.