⚠ **Try again once you are ready**

Grade received 70%   **Latest Submission Grade** 70%   **To pass** 80% or higher

Retake the assignment in **23h 47m**

---

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.

   ◉ 60% train. 20% dev. 20% test.

   ○ 98% train. 1% dev. 1% test.

   [↗ Expand]

   ✓ **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. The dev and test set should:

1 / 1 point

○ Be identical to each other (same (x,y) pairs)

⦿ Come from the same distribution

○ Come from different distributions

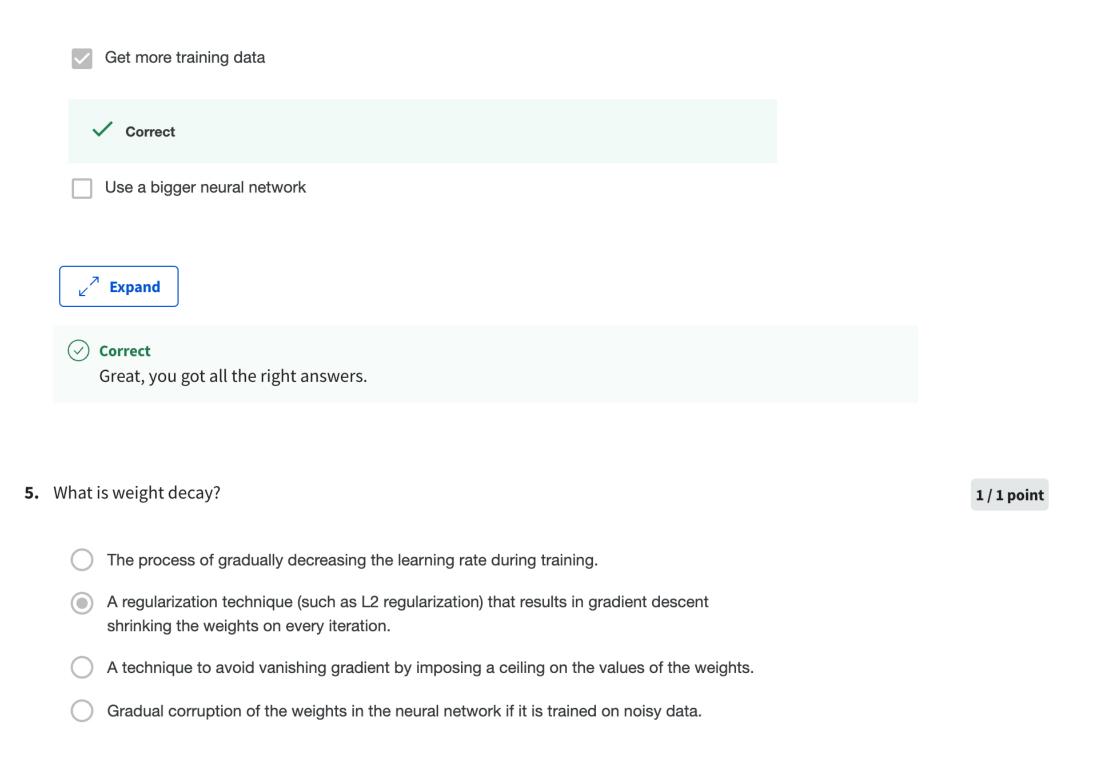○ Have the same number of examples

[⤢ **Expand**]

✓ **Correct**

3. A model developed for a project is presenting high bias. One of the sponsors of the project offers some resources that might help reduce the bias. Which of the following additional resources has a better chance to help reduce the bias?
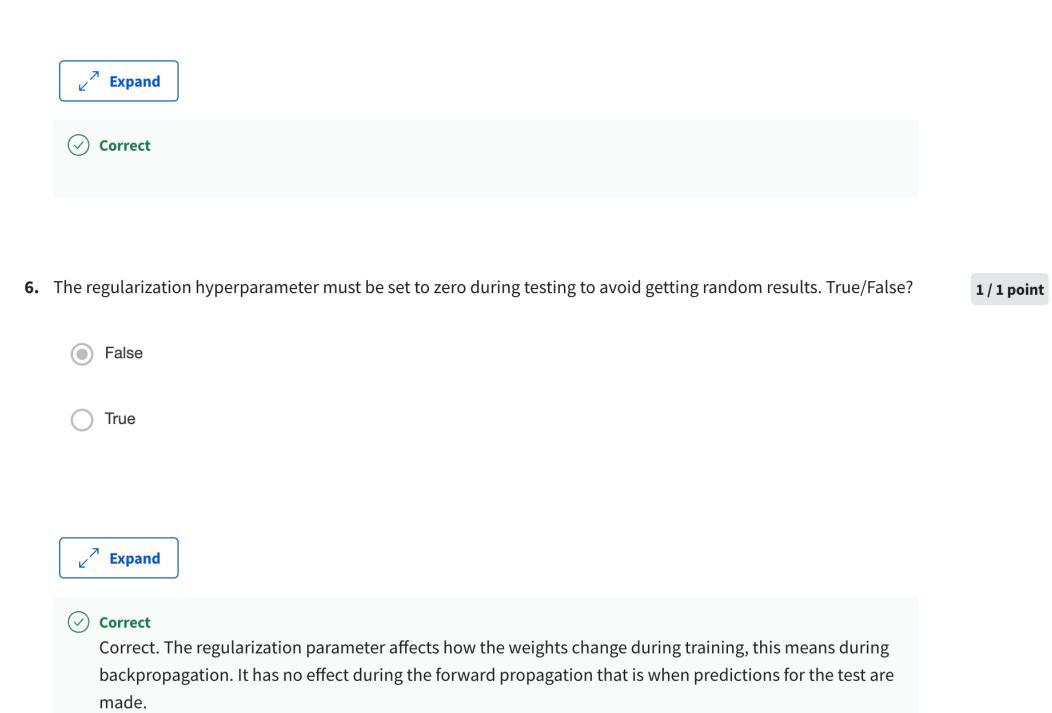
0 / 1 point

○ Gather more data for the project.

○ Give access to more computational resources like GPUs.

◉ Use different sources to gather data and better test the model.

[↗ Expand]

⊗ **Incorrect**
   No. More test data won't help reduce the bias.

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

☐ Use a bigger neural network

⤢ **Expand**

5. What is weight decay?   **1 / 1 point**

○ The process of gradually decreasing the learning rate during training.

◉ 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.

○ Gradual corruption of the weights in the neural network if it is trained on noisy data.

6. The regularization hyperparameter must be set to zero during testing to avoid getting random results. True/False?   **1 / 1 point**

- ⦿ False

- ◯ True

✓ **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. 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.

☐ 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.

⤢ **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?

○ Use Xavier initialization.

○ Use a larger regularization parameter.

○ Increase the number of layers of the network.

◉ Increase the number of cycles during the training.

⤢ **Expand**

⊗ **Incorrect**
    Incorrect. If the gradient is almost zero, more iterations won't make significant changes to the parameters.

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

☐ Decrease the value of keep_prob in dropout.

☐ Decrease the value of keep_prob in dropout.

✓ Decrease the value of the hyperparameter lambda.

> **!** **This should not be selected**
>
> Incorrect. When increasing the hyperparameter lambda, we increase the effect of the L_2 penalization.

☐ Increase the value of keep_prob in dropout.

☐ Use Xavier initialization.

✓ Increase the value of the hyperparameter lambda.

> ✓ **Correct**
>
> Correct. When increasing the hyperparameter lambda, we increase the effect of the L_2 penalization.

[↗ **Expand**]

⊗ **Incorrect**
You didn't select all the correct 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?

○ It will increase the variance of the model.

◉ It will make the training faster.

○ It will make the data easier to visualize.

○ 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.