1.Question 1

**If you have 10,000,000 examples, how would you split the train/dev/test set?**

98% train . 1% dev . 1% test

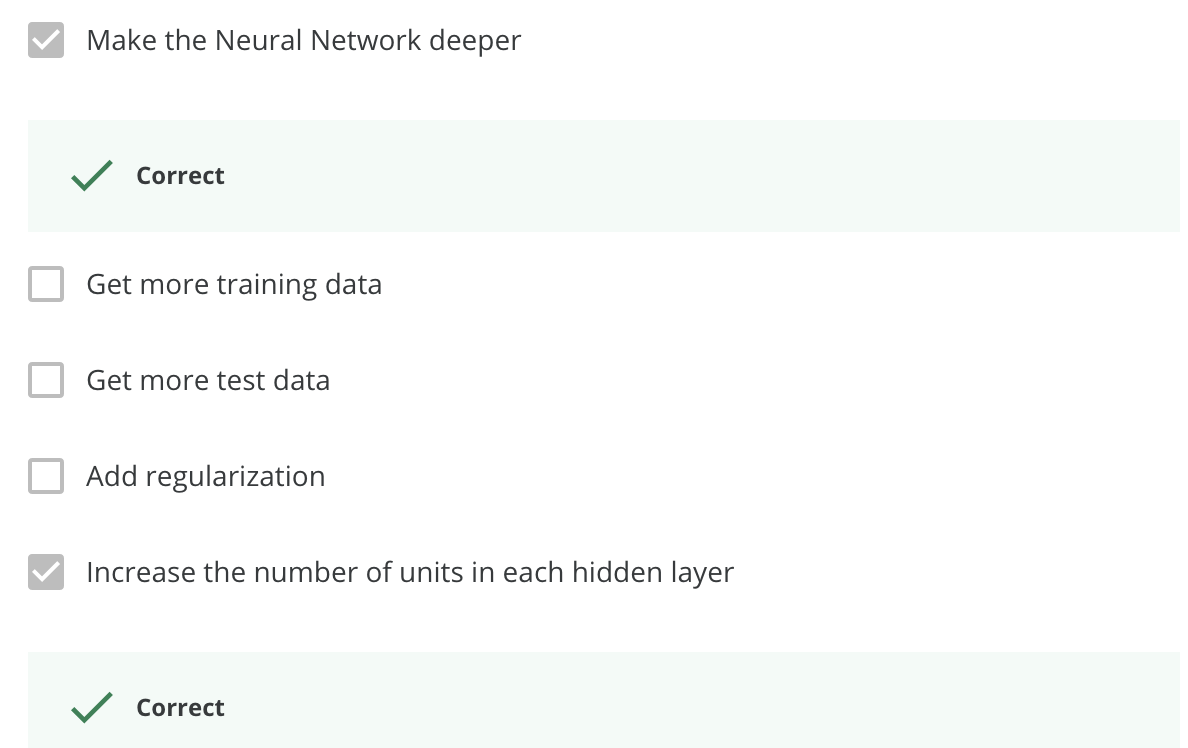
**2.Question 2**

**The dev and test set should:**

Come from the same distribution

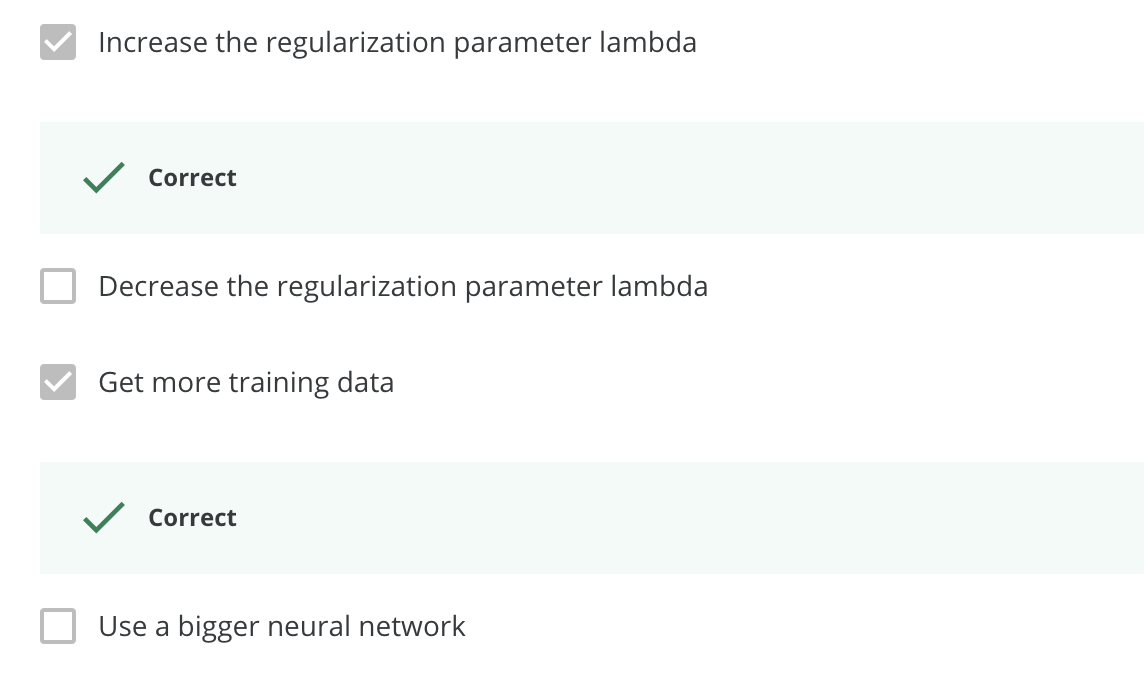
**3.Question 3**

**If your Neural Network model seems to have high bias, what of the following would be promising things to try? (Check all that apply.)**



**4.Question 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.)**



**5.Question 5**

**What is weight decay?**

A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration

**6.Question 6**

**What happens when you increase the regularization hyperparameter lambda?**

Weights are pushed toward becoming smaller(close to 0)

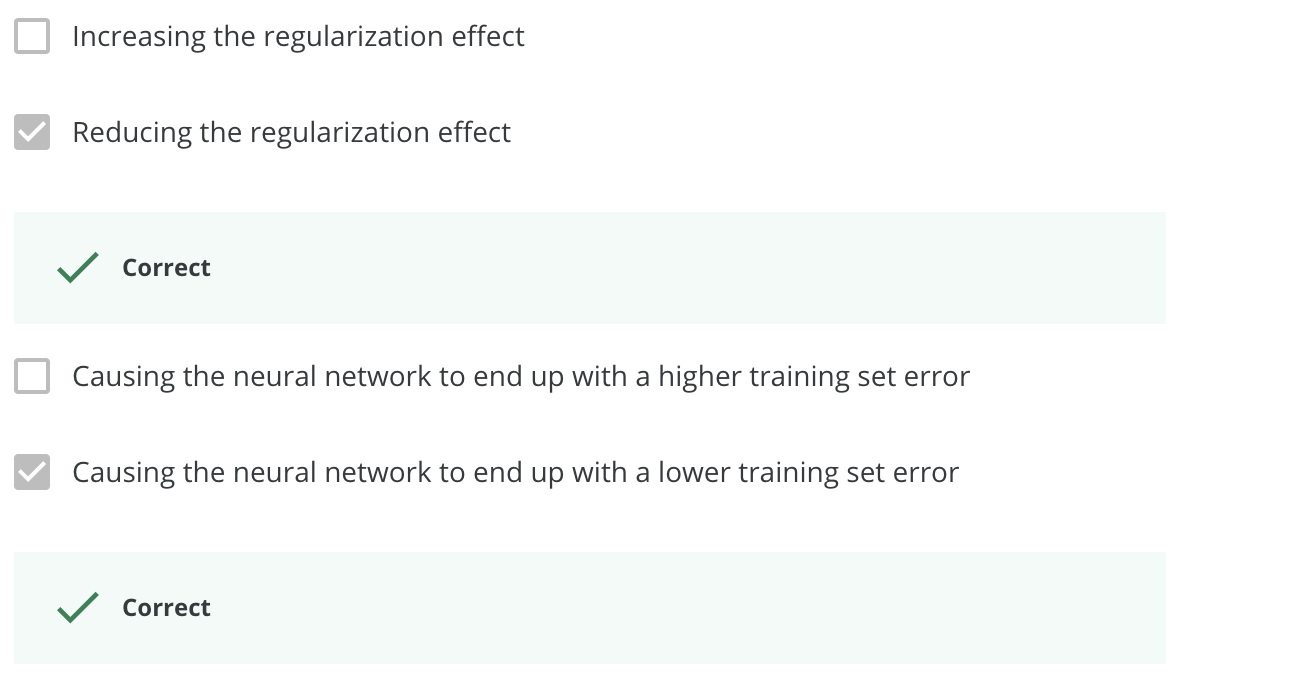
**7.Question 7**

**With the inverted dropout technique, at test time:**

You do not apply dropout (randomly eliminating units) and do not keep the 1/keep\_prob factor in the calculations used in training.

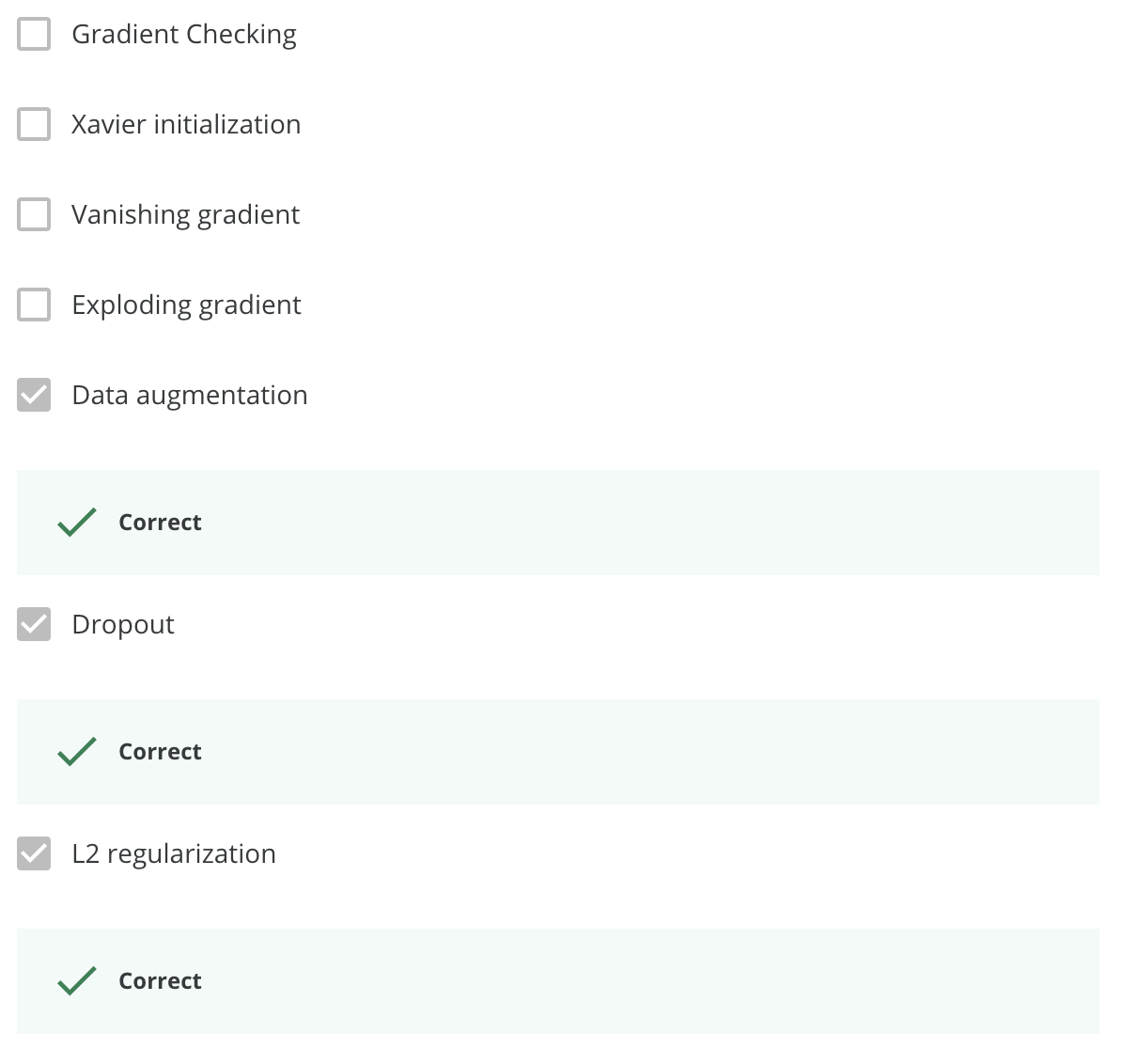
**8.Question 8**

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



**9.Question 9**

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



**10.Question 10**

**Why do we normalize the inputs x*x*?**

It makes the cost function faster to optimize.