1. If you have 10,000 examples, how would you split the train/dev/test set? Choose the best option. 1 point

- ○ 33% train. 33% dev. 33% test.
- ◉ 60% train. 20% dev. 20% test.
- ○ 98% train. 1% dev. 1% test.

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 point

- ○ He won't be able to measure the bias of the model.
- ○ Not having a test set is unacceptable under any circumstance.
- ◉ He might be overfitting to the dev set.
- ○ He won't be able to measure the variance of the model.

3. If your Neural Network model seems to have high variance, what of the following would be promising things to try? 1 point

- ☑ Add regularization
- ☑ Get more test data
- ☐ Increase the number of units in each hidden layer
- ☐ Make the Neural Network deeper
- ☐ Get more training data

4. You are working on an automated check-out kiosk for a supermarket, and are building a classifier for apples, bananas, and oranges. 1 point

Suppose your classifier obtains a training set error of 0.5%, and a development set error of 7%.

Which of the following strategies are most likely to reduce the development set error? (Check all that apply.)

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

5. What is weight decay? 1 point

- ○ Gradual corruption of the weights in the neural network if it is trained on noisy data.
- ○ 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.
- ○ The process of gradually decreasing the learning rate during training.

6. **True or False:** In L2 regularization, the lambda hyperparameter directly influences the calculations used by the model to make predictions during testing. 1 point

- ○ True
- ◉ False

7. With the inverted dropout technique, at test time: 1 point

- ◉ You do not apply dropout (do not randomly eliminate units) and do not keep the 1/keep_prob factor in the calculations used in training
- ○ You do not apply dropout (do not randomly eliminate units), but keep the 1/keep_prob factor in the calculations used in training.
- ○ You apply dropout (randomly eliminating units) and do not keep the 1/keep_prob factor in the calculations used in training
- ○ You apply dropout (randomly eliminating units) but keep the 1/keep_prob factor in the calculations used in training.

8. Decreasing the parameter keep_prob from (say) 0.6 to 0.4 will likely cause the following: 1 point

- ○ Causing the neural network to have a higher variance.
- ○ Reducing the regularization effect.
- ◉ Increasing the regularization effect.

9. Which of these techniques are useful for reducing variance (reducing overfitting)? (Check all that apply.) 1 point

- ☑ Dropout
- ☐ Xavier initialization
- ☑ Data augmentation
- ☑ L2 regularization
- ☐ Gradient Checking
- ☐ Exploding gradient
- ☐ Vanishing gradient

10. Why do we normalize the inputs x? 1 point

- ○ It makes it easier to visualize the data
- ○ It makes the parameter initialization faster
- ◉ It makes the cost function faster to optimize
- ○ Normalization is another word for regularization--It helps to reduce variance