✓ **Congratulations! You passed!**

Grade received 90%   **Latest Submission Grade** 90%   **To pass** 80% or higher   [**Go to next item**]

---

1. If you have 20,000,000 examples, how would you split the train/dev/test set? Choose the best option.   **1 / 1 point**

   ○ 60% train. 20% dev. 20% test.

   ○ 90% train. 5% dev. 5% test.

   ◉ 99% train. 0.5% dev. 0.5% test.

   [⤢ **Expand**]

   ✓ **Correct**
   Yes. Given the size of the dataset, 0.5% of the samples are enough to get a good estimate of how well the model is doing.

2. When designing a neural network to detect if a house cat is present in the picture, 500,000 pictures of cats were taken by their owners. **These are used to make the training, dev and test sets.** It is decided that to increase the size of the test set, 10,000 new images of cats taken from security cameras are going to be used in the test set. Which of the following is true?   **1 / 1 point**

   ○ This will increase the bias of the model so the new images shouldn't be used.

   ○ This will reduce the bias of the model and help improve it.

   ◉ This will be harmful to the project since now dev and test sets have different distributions.

   [⤢ **Expand**]

   ✓ **Correct**
   Yes. The quality and type of images are quite different thus we can't consider that the dev and the test sets came from the same distribution.

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

   ☐ Make the Neural Network deeper

   ☑ Get more training data

✓ **Correct**

   ☐ Increase the number of units in each hidden layer

   ☐ Get more test data

   ☑ Add regularization

✓ **Correct**

   [⤢ **Expand**]

   ✓ **Correct**
   Great, you got all the right answers.

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 19% and a dev set error of 21%. Which of the following are promising things to try to improve your classifier? (Check all that apply, suppose the human error is approximately 0%)   **1 / 1 point**

   ○ Get more training data.

   ○ Increase the regularization parameter lambda.

○ Use a bigger network.

[↗ Expand]

✓ **Correct**
Yes. This can be helpful to reduce the bias of the model, and then we can start trying to reduce the high variance if this happens.

---

5. What is weight decay?   0 / 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.

[↗ Expand]

✗ **Incorrect**

---

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

○ True

◉ False

[↗ Expand]

✓ **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?   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.

☑ It helps to reduce overfitting.

> ✓ **Correct**
> Correct. The dropout is a regularization technique and thus helps to reduce the overfit.

☑ In practice, it eliminates units of each layer with a probability of 1- keep_prob.

> ✓ **Correct**
> Correct. The probability that dropout doesn't eliminate a neuron is keep_prob.

[↗ 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?   1 / 1 point

○ Use a larger regularization parameter.

○ Increase the number of cycles during the training.

◉ Use Xavier initialization.

○ Increase the number of layers of the network.

✓ **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**

☐ Xavier initialization

☐ Vanishing gradient

☑ Data augmentation

> ✓ **Correct**

☑ Dropout

> ✓ **Correct**

☑ L2 regularization

> ✓ **Correct**

☐ Exploding gradient

☐ Gradient Checking

✓ **Correct**
Great, you got all the right 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?   **1 / 1 point**

○ It will increase the variance of the model.

○ It won't have any positive or negative effects.

◉ It will make the training faster.

○ It will make the data easier to visualize.

✓ **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.