Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to 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**

- 98% train. 1% dev. 1% test.
- 33% train. 33% dev. 33% test.
- (●) 60% train. 20% dev. 20% 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**

- Have the same number of examples
- (●) Come from the same distribution
- Be identical to each other (same (x,y) pairs)
- Come from different distributions

Expand

Correct

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.) **1 / 1 point**

- [x] Make the Neural Network deeper

Correct

- [ ] Get more training data
- [x] Increase the number of units in each hidden layer

Correct

- [ ] Add regularization

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**

- Increase the regularization parameter lambda.
- Get more training data.
- 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? **1 / 1 point**

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

Expand

Correct

6. What happens when you increase the regularization hyperparameter lambda? **1 / 1 point**

- Gradient descent taking bigger steps with each iteration (proportional to lambda)
- Weights are pushed toward becoming bigger (further from 0)
- Weights are pushed toward becoming smaller (closer to 0)
- Doubling lambda should roughly result in doubling the weights

Doubling lambda should roughly result in doubling the weights

Expand

Correct

7. Which of the following are true about dropout? **1 / 1 point**

- [ ] It helps to reduce the bias of a model.
- [x] 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 keep_prob.
- [x] 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**

- ( ) Increase the number of layers of the network.
- ( ) Use a larger regularization parameter.
- (•) Use Xavier initialization.
- ( ) Increase the number of cycles during the training.

Expand

Correct
Correct. A careful initialization can help reduce the vanishing gradient problem.

9. Which of the following actions increase the regularization of a model? (Check all that apply) **1 / 1 point**

- [ ] Increase the value of keep_prob in dropout.

- [x] Increase the value of the hyperparameter lambda.

Correct
Correct. When increasing the hyperparameter lambda we increase the effect of the L_2 penalization.

- [ ] Decrease the value of the hyperparameter lambda.
- [x] Make use of data augmentation.

Correct
Correct. Data augmentation has a way to generate "new" data at a relatively low cost. Thus making use of data augmentation can reduce the variance.

- [ ] Normalizing the data.

Expand

Correct
Great, you got all the right answers.

10. Which of the following is the correct expression to normalize the input $\mathbf{x}$?

1 / 1 point

- (●) $x = \frac{x - \mu}{\sigma}$
- ( ) $x = \frac{x}{\sigma}$
- ( ) $x = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$
- ( ) $x = \frac{1}{m}\sum_{i=1}^{m} (x^{(i)})^2$

Expand

Correct
Correct. This shifts the mean of the input to the origin and makes the variance one in each coordinate of the input examples.