Congratulations! You passed!

Grade received 91% To pass 80% or higher

Go to next item

Practical aspects of Deep Learning

Latest Submission Grade 91%

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.

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

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**

- ☐ Add regularization
- ☐ Get more training data
- ☑ Increase the number of units in each hidden layer

Correct

- ☑ Make the Neural Network deeper

Correct

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

Correct

- ☐ Use a bigger neural network

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

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

Correct

6. To reduce high variance, the regularization hyperparameter lambda must be increased. True/False? **1 / 1 point**

- ○ False
- ◉ True

Correct
Correct. By increasing the regularization parameter the magnitude of the weight parameters is reduced. This helps reduce the variance.

7. Which of the following are true about dropout? **0.5 / 1 point**

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

- [x] In practice, it eliminates units of each layer with a probability of keep_prob.

This should not be selected
Incorrect. The probability that dropout doesn't eliminate a neuron is keep_prob.

8. Increasing the parameter keep_prob from (say) 0.5 to 0.6 will likely cause the following: (Check the two that apply) **1 / 1 point**

- [ ] Increasing the regularization effect
- [x] Reducing the regularization effect

Correct

- [ ] Causing the neural network to end up with a higher training set error
- [x] Causing the neural network to end up with a lower training set error

Correct

9. Which of the following actions increase the regularization of a model? (Check all that apply) **0.6 / 1 point**

- [ ] Decrease the value of the hyperparameter lambda.
- [ ] Decrease the value of keep_prob in dropout.
- [ ] Increase the value of keep_prob in dropout.
- [ ] Use Xavier initialization.
- [ ] Increase the value of the hyperparameter lambda.

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? **1 / 1 point**

- ( ) It will make the data easier to visualize.
- ( ) It will increase the variance of the model.
- ( ) It won't have any positive or negative effects.
- (•) It will make the training faster.

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.