Practical aspects of deep learning

LATEST SUBMISSION GRADE

100%

---

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

   ○ 33% train . 33% dev . 33% test

   ● 98% train . 1% dev . 1% test

   ○ 60% train . 20% dev . 20% test

   ✓ **Correct**

2. The dev and test set should: `1 / 1 point`

   ● Come from the same distribution

   ○ Come from different distributions

   ○ Be identical to each other (same (x,y) pairs)

   ○ **Have the same number of examples**

   ✓ **Correct**

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

☐ Increase the number of units in each hidden layer

☑ Get more training data

> ✓ **Correct**

☑ Add regularization

> ✓ **Correct**

☐ Make the Neural Network deeper

☐ Get more test data

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?

○ 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.

○ 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.

✓ Correct

6. What happens when you increase the regularization hyperparameter lambda?

⦿ Weights are pushed toward becoming smaller (closer to 0)

○ Weights are pushed toward becoming bigger (further from 0)

○ Doubling lambda should roughly result in doubling the weights

○ Gradient descent taking bigger steps with each iteration (proportional to lambda)

✓ Correct

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

◯ You apply dropout (randomly eliminating 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 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.

✓ **Correct**

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

□ Increasing the regularization effect

☑ Reducing the regularization effect

✓ **Correct**

□ Causing the neural network to end up with a higher training set error

☑ Causing the neural network to end up with a lower training set error

✓ **Correct**

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

- [x] L2 regularization

  ✓ **Correct**

- [ ] Vanishing gradient

- [x] Data augmentation

  ✓ **Correct**

- [x] Dropout

  ✓ **Correct**

- [ ] Exploding gradient

- [ ] Xavier initialization

- [ ] Gradient Checking

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

- ( ) It makes the parameter initialization faster

- ( ) It makes it easier to visualize the data

- (•) It makes the cost function faster to optimize

- ( ) Normalization is another word for regularization--It helps to reduce variance

  ✓ **Correct**