

Practical aspects of deep learning

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

- ☐ 33% train . 33% dev . 33% test
 - ☒ 98% train . 1% dev . 1% test
 - ☐ 60% train . 20% dev . 20% test
-

2. The dev and test set should:

- ☐ Come from different distribution.
 - ☒ Come from the same distribution.
 - ☐ **Have the same number of examples.**
-

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

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

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

5. What is weight decay?

- ☐ 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 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.
-

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).
-

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

- ☐ You do not apply dropout (do not randomly eliminate units), but keep the $1/\text{keep_prob}$ factor in the calculations used in training.
 - ☐ You apply dropout (randomly eliminate units) but keep the $1/\text{keep_prob}$ factor in the calculations used in training.
 - ☐ You apply dropout (randomly eliminate units) and do not keep the $1/\text{keep_prob}$ factor in the calculations used in training.
 - ☒ You do not apply dropout (do not randomly eliminate units) and do not keep the $1/\text{keep_prob}$ factor in the calculations used in training.
-

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.
 - ☐ 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.
-

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

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

10. Why do we normalize the inputs x?

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

