Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization

Week 1: Practical aspects of Deep Learning

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

Answer: 98 train, 1 dev, 1 test.

Comment: See lecture notes. Answer will be different if there are 10,000 examples.

2 The dev and test set should:

Answer: Come 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?

Answer: Get more training data; Add regularization.

Comment: See lecture notes. Note that if high bias, then make the neural network deeper, and increase the number of units in each hidden layer.

4 Version 1: Working on a model to classify bananas and oranges your classifier gets a training set error of 0.1 and a dev set error of 11. Which of the following two are true?

Answer: The model has a high variance; The model is overfitting the train set.

Comment: See lecture notes.

4 Version 2: 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)

Answer: Increase the regularization parameter lambda; Get more training data.

5 Version 1: In every case it is a good practice to use dropout when training a deep neural network because it can help to prevent overfitting. True/False?

Answer: False.

Comment: In most cases, it is recommended to not use dropout if there is no overfit. Although in computer vision, due to the nature of the data, it is the default practice.

5 Version 2: What is weight decay?

Answer: A regularization technique (such as L2 regularization) that results in gradient descent shrinking the weights on every iteration.

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

Answer: False.

Comment: 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.

6 Version 2: What happens when you increase the regularization hyperparameter lambda?

Answer: Weights are pushed toward becoming smaller (closer to 0).

7 With the inverted dropout technique, at test time:

Answer: 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.

8 Version 1: 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?

Answer: Use Xavier initialization.

Comment: A careful initialization can help reduce the vanishing gradient problem.

8 Version 2: Increasing the parameter keep prob from 0.5 to 0.6 will likely cause the following: (Check the two that apply)

Answer: Reducing the regularization effect; 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.)

Answer: Dropout, L2 regularization, Data augmentation.

10 Version 1: 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?

Answer: It will make the training faster.

Comment: 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.

10 Version 2: Why do we normalize the inputs x?

Answer: It makes the cost function faster to optimize.