

Practical aspects of deep learning

测验, 10 个问题

第 1 个问题

1

point

1。第 1 个问题

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

98% train . 1% dev . 1% test

33% train . 33% dev . 33% test

60% train . 20% dev . 20% test

第 2 个问题

1

point

2。第 2 个问题

The dev and test set should:

Come from the same distribution

Come from different distributions

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

Have the same number of examples

第 3 个问题

1

point

3。第 3 个问题

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

Add regularization

Make the Neural Network deeper

Increase the number of units in each hidden layer

Get more test data

Get more training data

第 4 个问题

1

point

4。第 4 个问题

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 λ

Decrease the regularization parameter λ

Get more training data

Use a bigger neural network

第 5 个问题

1

point

5。第 5 个问题

What is weight decay?

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.

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.

第 6 个问题

1

point

6。第 6 个问题

What happens when you increase the regularization hyperparameter λ ?

Weights are pushed toward becoming smaller (closer to 0)

Weights are pushed toward becoming bigger (further from 0)

Doubling λ should roughly result in doubling the weights

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

第 7 个问题

1

point

7。第 7 个问题

With the inverted dropout technique, at test time:

You apply dropout (randomly eliminating 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

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 eliminating units) but keep the $1/\text{keep_prob}$ factor in the calculations used in training.

第 8 个问题

1

point

8。第 8 个问题

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

第 9 个问题

1

point

9。第 9 个问题

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

Xavier initialization

Gradient Checking

Exploding gradient

Vanishing gradient

Dropout

L2 regularization

Data augmentation

第 10 个问题

1

point

10。第 10 个问题

Why do we normalize the inputs x ?

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

It makes it easier to visualize the data

It makes the parameter initialization faster

It makes the cost function faster to optimize