

1. Which of the following statements is true?

The deeper layers of a neural network are typically computing more complex features of the input than the earlier layers.

2. Considering the following statement:

Vectorization allows you to compute forward propagation in an L-layer neural network without an explicit for-loop (or any other explicit iterative loop) over the layers $l=1, 2, \dots, L$.

False

3. What does a neuron compute?

A neuron computes a linear function ($z = Wx + b$) followed by an activation function

4. Suppose you have built a neural network. You decide to initialize the weights and biases to be zero. Which of the following statement is true?

Each neuron in the first hidden layer will perform the same computation. So even after multiple iterations of gradient descent each neuron in the layer will be computing the same thing as other neurons.

5. Considering the following statement:

Logistic regression's weights w should be initialized randomly rather than to all zeros, because if you initialize to all zeros, then logistic regression will fail to learn a useful decision boundary because it will fail to "break symmetry".

False

6. Considering the following statement:

During forward propagation, in the forward function for a layer l you need to know what is the activation function in a layer (Sigmoid, tanh, ReLU, etc.). During backpropagation, the

corresponding backward function also needs to know what is the activation function for layer l , since the gradient depends on it.

True

7. Considering the following statement:

When a decision tree is grown to full depth, it is more likely to fit the noise in the data.

True

8. Considering the following statement:

When the hypothesis space is richer, over fitting is more likely.

True

9. Considering the following statement:

When the feature space is larger, over fitting is more likely.

True

10. Suppose you have picked the parameter θ for a model using 10-fold cross validation(CV).
The best way to pick a final model to use and estimate its error is to

Train a new model on the full data set, using the θ you found; use the average CV error as its error estimate

11. Suppose we want to compute 10-Fold Cross-Validation error on 100 training examples. We need to compute error N_1 times, and the Cross-Validation error is the average of the errors. To compute each error, we need to build a model with data of size N_2 , and test the model on the data of size N_3 .

What are the appropriate numbers for N1, N2, N3?

N1 = 10, N2 = 90, N3 = 10

12. MLE estimates are often undesirable because

They are biased

13. Which of the following tends to work best on small data sets (few observations)?

Naive Bayes

14. Which of the following regularization method(s) is(are) scale-invariant?

L0 but not L1 or L2

15. Consider the following confusion matrix

		Current Answer	
		True	False
Predicted Answer	True	8	2
	False	12	11

For the above “confusion matrix” the precision is

None of the above