Quiz 6

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.
○	The earlier layers of a neural network are typically computing more complex features of the input than the deeper 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, …,L.**

○	True
◉	False

3.
What does a neuron compute?

○	A neuron computes an activation function followed by a linear function (z = Wx + b)
◉	A neuron computes a linear function (z = Wx + b) followed by an activation function
○	A neuron computes the mean of all features before applying the output to an activation function
○	A neuron computes a function g that scales the input x linearly (Wx + b)

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 in the first iteration. But after one iteration of gradient descent they will learn to compute different things because we have “broken symmetry”.

(selected)	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.
	Each neuron in the first hidden layer will compute the same thing, but neurons in different layers will compute different things, thus we have accomplished “symmetry breaking” as described in lecture.
	The first hidden layer’s neurons will perform different computations from each other even in the first iteration; their parameters will thus keep evolving in their own way.

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”.**

	True
(selected)	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.**

(selected)	True
	False

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.**

(selected)	True
	False

8.
Considering the following statement:

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

◉	True
○	False

9.
Considering the following statement:

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

◉	True
○	False

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

○	Pick any of the 10 models you built for your model; use its error estimate on the held-out data
○	Pick any of the 10 models you built for your model; use the average CV error for the 10 models as its error estimate
◉	Average all of the 10 models you got; use the average CV error as its error estimate
○	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 N1 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 N2, and test the model on the data of size N3.

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

◉	N1 = 10, N2 = 90, N3 = 10
○	N1 = 1, N2 = 90, N3 = 10
○	N1 = 10, N2 = 100, N3 = 10

○	N1 = 10, N2 = 100, N3 = 100

12.
MLE estimates are often undesirable because

○	They are biased
◉	They have high variance
○	They are not consistent estimators
○	None of the above

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

◉	Naive Bayes
○	Logistic regression

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

○	L0 and L2 but not L1
◉	L1 and L2 but not L0
○	L0 but not L1 or L2
○	L2 but not L0 or L1

15.
Consider the following confusion matrix

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

For the above "confusion matrix" the precision is

○	2/10
◉	8/20
○	19/33
○	None of the above