Suppose D1 is a 10x6 matrix and D2 is a 1x11 matrix. You set:

DVec = [D1(:); D2(:)];

Which of the following would get D2 back from DVec?

- reshape(DVec(60:71), 1, 11)
- reshape(DVec(61:72), 1, 11)
- reshape(DVec(61:71), 1, 11)

Correct

- reshape(DVec(60:70), 11, 1)

Continue

Let $J(\theta) = \theta^3$. Furthermore, let $\theta = 1$ and $\epsilon = 0.01$. You use the formula:

$$\frac{J(\theta+\epsilon)-J(\theta-\epsilon)}{2\epsilon}$$

to approximate the derivative. What value do you get using this approximation? (When $\theta = 1$, the true, exact derivative is $\frac{d}{d\theta}J(\theta) = 3$).

- 3.0000
- 3.0001

Correct

- 3.0301
- 6.0002

Continue

What is the main reason that we use the backpropagation algorithm rather than the numerical gradient computation method during learning?

- The numerical gradient computation method is much harder to implement.
- The numerical gradient algorithm is very slow.

Correct

- Backpropagation does not require setting the parameter EPSILON.
- None of the above.

Consider this procedure for initializing the parameters of a neural network:

1. Pick a random number r = rand(1,1) * (2 * INIT_EPSILON) - INIT_EPSILON;
2. Set $\Theta_{ij}^{(l)} = r$ for all i, j, l.

Does this work?

- Yes, because the parameters are chosen randomly.
- Yes, unless we are unlucky and get r=0 (up to numerical precision).
- Maybe, depending on the training set inputs x(i).
- No, because this fails to break symmetry.

Correct

Continue

Suppose you are using gradient descent together with backpropagation to try to minimize $J(\Theta)$ as a function of Θ. Which of the following would be a useful step for verifying that the learning algorithm is running correctly?

- Plot $J(\Theta)$ as a function of Θ, to make sure gradient descent is going downhill.
- Plot $J(\Theta)$ as a function of the number of iterations and make sure it is increasing (or at least non-decreasing) with every iteration.
- Plot $J(\Theta)$ as a function of the number of iterations and make sure it is decreasing (or at least non-increasing) with every iteration.

  Correct

- Plot $J(\Theta)$ as a function of the number of iterations to make sure the parameter values are improving in classification accuracy.