1. Which of the following are true about hyperparameter search? 1 point

- ○ When using random values for the hyperparameters they must be always uniformly distributed.
- ◉ Choosing random values for the hyperparameters is convenient since we might not know in advance which hyperparameters are more important for the problem at hand.
- ○ Choosing values in a grid for the hyperparameters is better when the number of hyperparameters to tune is high since it provides a more ordered way to search.
- ○ When sampling from a grid, the number of values for each hyperparameter is larger than when using random values.

2. In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply. 1 point

- ☑ mini-batch size
- ☐ β_1, β_2 in Adam.
- ☑ α
- ☐ ϵ in Adam.
- ☑ The β parameter of the momentum in gradient descent.

3. Using the "Panda" strategy, it is possible to create several models. True/False? 1 point

- ○ False
- ◉ True

4. Knowing that the hyperparameter α should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for α? 1 point

- ○
```
r = 4*np.random.rand()
alpha = 10**r
```
- ○
```
r = -5*np.random.rand()
alpha = 10**r
```
- ○
```
r = np.random.rand()
alpha = 0.001 + r*0.999
```
- ◉
```
r = -3*np.random.rand()
alpha = 10**r
```

5. Finding good hyperparameter values is very time-consuming. So typically you should do it once at the start of the project, and try to find very good hyperparameters so that you don't ever have to tune them again. True or false? 1 point

- ◉ False
- ○ True

6. When using batch normalization it is OK to drop the parameter $W^{[l]}$ from the forward propagation since it will be subtracted out when we compute $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False? 1 point

- ○ True
- ◉ False

7. When using normalization: 1 point

$z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$

In case σ is too small, the normalization of $z^{(i)}$ may fail since division by 0 may be produced due to rounding errors. True/False?

- ◉ False
- ○ True

8. Which of the following is true about batch normalization? 1 point

- ○ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.
- ○ $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- ◉ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\tilde{z}^{[l]}$.
- ○ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.

9. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should: 1 point

- ○ Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.
- ○ Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.
- ○ If you implemented Batch Norm on mini-batches of (say) 256 examples, then to evaluate on one test example, duplicate that example 256 times so that you're working with a mini-batch the same size as during training.
- ◉ Perform the needed normalizations, use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training.

10. Which of these statements about deep learning programming frameworks are true? (Check all that apply) 1 point

- [ ] Deep learning programming frameworks require cloud-based machines to run.
- [x] A programming framework allows you to code up deep learning algorithms with typically fewer lines of code than a lower-level language such as Python.
- [x] Even if a project is currently open source, good governance of the project helps ensure that it remains open even in the long term, rather than become closed or modified to benefit only one company.