With a relatively small set of hyperparameters, it is OK to use a grid search. True/False? 1 point

- ◯ True
- ◯ False

2.If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose? 1 point

- ☐ The $\beta$ parameter of the momentum in gradient descent.
- ☐ $\alpha$
- ☐ $\epsilon$ in Adam.
- ☐ $\beta_1, \beta_2$ in Adam.

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

- ◯ False
- ◯ True

4.If you think $\beta$ (hyperparameter for momentum) is between 0.9 and 0.99, which of the following is the recommended way to sample a value for beta? 1 point

- ◯ r = np.random.rand() beta = r*0.9 + 0.09
- ◯ r = np.random.rand() beta = 1-10**(- r + 1)
- ◯ r = np.random.rand() beta = r*0.09 + 0.9
- ◯ r = np.random.rand() beta = 1-10**(- r - 1)

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.In batch normalization as presented in the videos, if you apply it on the $l$th layer of your neural network, what are you normalizing? 1 point

- ◯ $z^{[l]}$
- ◯ $W^{[l]}$
- ◯ $b^{[l]}$
- ◯ $a^{[l]}$

7.In the normalization formula $z^{(i)}_{norm} = \frac{z^{(i)}-\mu}{\sqrt{\sigma^2+\varepsilon}}$, why do we use epsilon? 1 point

- ◯ To have a more accurate normalization

- ◯ In case $\mu$ is too small
- ◯ To speed up convergence
- ◯ To avoid division by zero

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

- ◯ The optimal values to use for $\gamma$ and $\beta$ 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.A neural network is trained with Batch Norm. At test time, to evaluate the neural network we turn off the Batch Norm to avoid random predictions from the network. True/False? 1 point

- ◯ False
- ◯ True

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

- ☐ 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.
- ☐ Deep learning programming frameworks require cloud-based machines to run.
- ☐ 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.