Hyperparameter tuning, Batch Normalization, Programming Frameworks

1.Question 1

If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False?



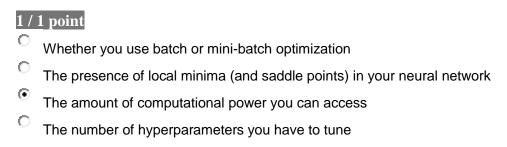
2.Question 2

Every hyperparameter, if set poorly, can have a huge negative impact on training, and so all hyperparameters are about equally important to tune well. True or False?



3.Question 3

During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by:



4.Question 4

If you think β (hyperparameter for momentum) is between on 0.9 and 0.99, which of the following is the recommended way to sample a value for beta?

```
1/1 point

r = np.random.rand()
beta = r*0.09 + 0.9
```

```
r = np.random.rand()

beta = 1-10**(- r - 1)

r = np.random.rand()

beta = 1-10**(- r + 1)

r = np.random.rand()

beta = r*0.9 + 0.09
```

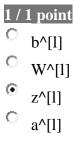
5.Question 5

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 revisit tuning them again. True or false?



6.Question 6

In batch normalization as presented in the videos, if you apply it on the llth layer of your neural network, what are you normalizing?



7.Question 7

In the normalization formula $z_{norm}^{(i)} = (z^{(i)} - \mu) / sqrt(\sigma^2 + \epsilon)$, why do we use epsilon?

```
    1 / 1 point
    In case μ is too small
    To have a more accurate normalization
    To avoid division by zero
    To speed up convergence
```

8.Question 8

Which of the following statements about γ and β in Batch Norm are true?

1/1	point
	There is one global value of $\gamma \in \Re$ and one global value of $\beta \in \Re$ for each layer, and applies to ne hidden units in that layer.
	They set the mean and variance of the linear variable $z^{\Lambda}[1]z[l]$ of a given layer.
	eta and $\S ammay$ are hyperparameters of the algorithm, which we tune via random sampling.
	They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just gradient descent.
	The optimal values are $\gamma = \operatorname{sqrt}(\sigma ^2 + \varepsilon)$, and $\beta = \mu$.
9.Question 9 After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should:	
1/1	point
weig one	Skip the step where you normalize using μ and σ^2 since a single test example cannot be nalized.
	Perform the needed normalizations, use μ and σ^2 estimated using an exponentially phted average across mini-batches seen during training.
	If you implemented Batch Norm on mini-batches of (say) 256 examples, then to evaluate on test example, duplicate that example 256 times so that you're working with a mini-batch the e size as during training.
	Use the most recent mini-batch's value of μ and σ ^2 to perform the needed normalizations.
10.Question 10 Which of these statements about deep learning programming frameworks are true? (Check all that apply)	
1/1	point
	Deep learning programming frameworks require cloud-based machines to run.
one	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 company.
✓ fewe	A programming framework allows you to code up deep learning algorithms with typically er lines of code than a lower-level language such as Python.