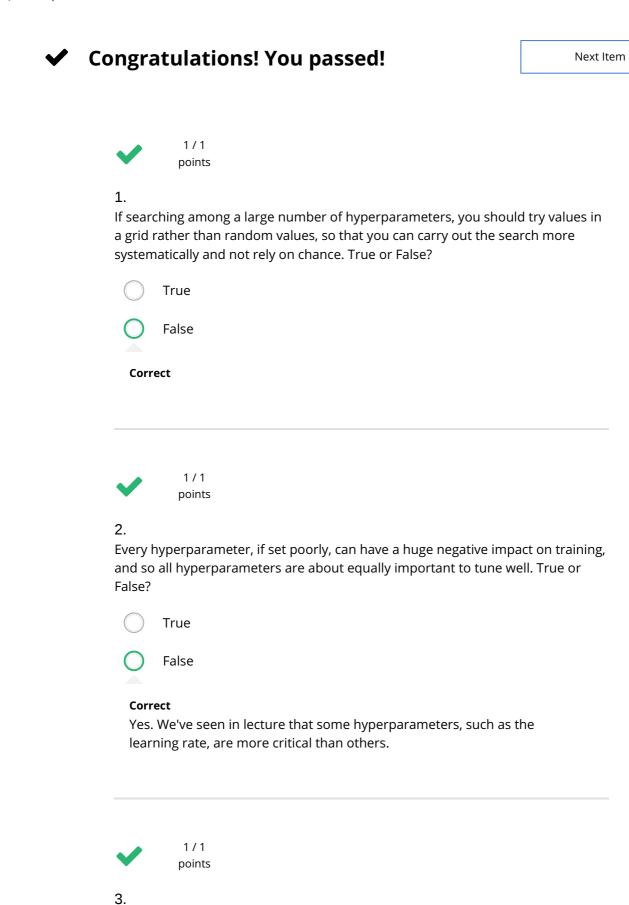
Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions



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:

Hyperparameter	tuning, Batch Normalization, Programming
Frameworks O	Whether you use batch or mini-batch optimization

10/10 points (100%)

Quiz, '	10 q	uestions
---------	------	----------

The presence of local minima (and saddle points) in your neural
network

The amount of computational power you can access

Correct

The number of hyperparameters you have to tune



1/1 points

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?

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



Correct

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

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

1/1 points

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

Hyperparameteratuming, Batch Normalization is Programming True Frameworks false?

10/10 points (100%)

Quiz,	10	aue	estic	ons

6/	- 10	
	- 1)	Irue
v.		HIUC



Correct



1/1 points

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



 $z^{[l]}$

Correct



 $b^{[l]}$



1/1 points

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

()	Tohavo	a mara	accurate	norma	lization
V //	TO HAVE	annore	accurate	поппа	1174111011

In case μ is too small

To speed up convergence

To avoid division by zero

Correct

Frameworl	meter tuning, Batch Normalization, Programming	10/10 points (100%)
Quiz, 10 questions	8.	
	Which of the following statements about γ and eta in Batch Norm are true?	
	$\hfill \beta$ and γ are hyperparameters of the algorithm, which we tune via random sampling.	
	Un-selected is correct	
	There is one global value of $\gamma\in\Re$ and one global value of $\beta\in\Re$ for each layer, and applies to all the hidden units in that layer.	
	Un-selected is correct	
	$igcap = 0$ The optimal values are $\gamma = \sqrt{\sigma^2 + arepsilon}$, and $eta = \mu$.	
	Un-selected is correct	
	They set the mean and variance of the linear variable $z^{[l]}$ of a given layer.	
	Correct	
	They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.	
	Correct	
	1/1 points	
	9. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should:	
	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 Hyperparameter tuning Batch Normalization; Programming me size as during training.

10/10 points (100%)

Quiz, 10 questions

Frameworks

0	Perform the needed normalizations, use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training.
Corre	ect
	Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.
~	1 / 1 points
	of these statements about deep learning programming frameworks are Check all that apply)
	Even if a project is currently open source, good governance of the project helps ensure that the it remains open even in the long term, rather than become closed or modified to benefit only one company.
Corre	ect
	Deep learning programming frameworks require cloud-based machines to run.
Un-se	elected is correct
	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.
Corre	ect



Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions