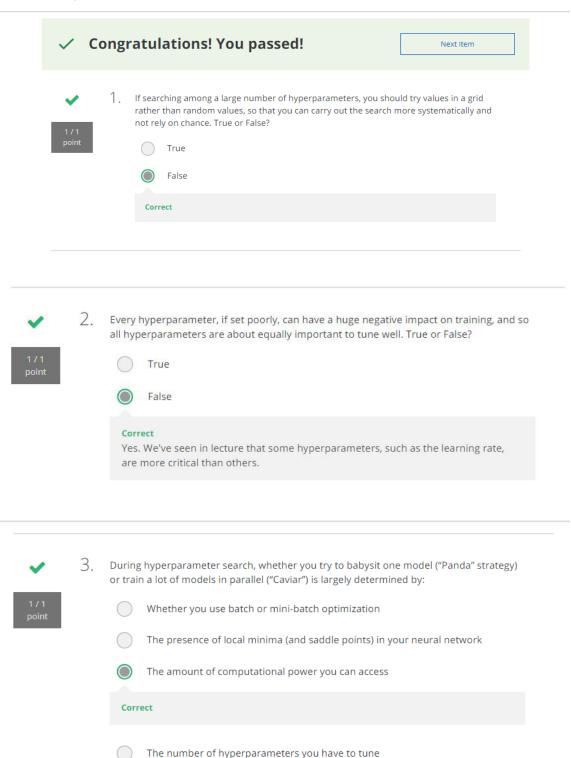
,	,		
۲	₹	•	

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Quiz, 10 questions

10/10 points (100%)



1/1 point	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
1/1 point	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? True False Correct
1/1 point	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? $ a^{[l]} $ $ W^{[l]} $ $ b^{[l]} $ $ z^{[l]} $ Correct

~	7. •	In the normalization formula $z_{norm}^{(i)}=rac{z^{(i)}-\mu}{\sqrt{\sigma^2+arepsilon}}$, why do we use epsilon?			
1/1 point		To avoid division by zero			
		Correct			
		To speed up convergence			
		On case μ is too small			
		To have a more accurate normalization			
~	8.	Which of the following statements about γ and eta in Batch Norm are true?			
1/1 point		$\hfill \beta$ and γ are hyperparameters of the algorithm, which we tune via random sampling.			
		Un-selected is correct			
		The optimal values are $\gamma=\sqrt{\sigma^2+arepsilon}$, and $\beta=\mu$.			
		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.					
Correc	t				
	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-sel	ected is co	rrect			

	~	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			Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.	
				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.	
				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.	
-	~	10.		of these statements about deep learning programming frameworks are true? all that apply)	
	1 / 1 point			Deep learning programming frameworks require cloud-based machines to run.	
			Un-s	elected is correct	
				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.	
			Corr	ect	
				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.	
			Corr	ect	