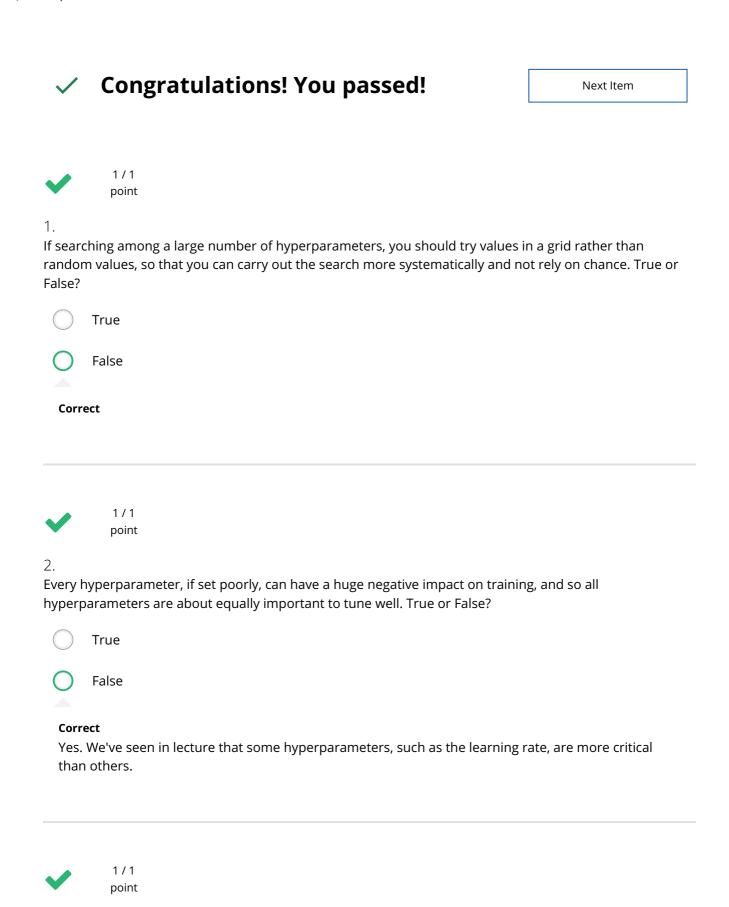
Hyperparameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions

3.



During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of

Hyperparameterctuning Rytch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questio Whether you use batch or mini-batch optimization

10 quest	lionister is a disconstruction with patern optimization	
	The presence of local minima (and saddle points) in your neural network	
0	The amount of computational power you can access	
Correct		
	The number of hyperparameters you have to tune	
✓ 4.	1/1 point	
If you t	think eta (hyperparameter for momentum) is between on 0.9 and 0.99, which of the following is the mended way to sample a value for beta?	
	1 r = np.random.rand() 2 beta = r*0.09 + 0.9	
0	1 r = np.random.rand() 2 beta = 1-10**(- r - 1)	
Corr	ect	
	1 r = np.random.rand() 2 beta = 1-10**(- r + 1)	
	1 r = np.random.rand() 2 beta = r*0.9 + 0.09	

Frafine w	arameter tuning, Batch Normalization, Programming 10/10 points 10/10 p
	True
0	False
Corr	ect
~	1 / 1 point
	h normalization as presented in the videos, if you apply it on the \emph{l} th layer of your neural network, re you normalizing?
0	$z^{[l]}$
Corr	ect
	$a^{[l]}$
	$b^{[l]}$
	$W^{[l]}$
~	1 / 1 point
7. In the	normalization formula $z_{norm}^{(i)}=rac{z^{(i)}-\mu}{\sqrt{\sigma^2+arepsilon}}$, why do we use epsilon?
	To have a more accurate normalization
	To speed up convergence

In case μ is too small

To avoid division by zero

Hypefffaffameter tuning, Batch Normalization, Programming Frameworks

10/10 points (100%)

Quiz, 10 questions

~	1 / 1 point	
8. Which	of the following statements about γ and eta in Batch Norm are true?	
	They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.	
Correct		
	They set the mean and variance of the linear variable $z^{[l]}$ of a given layer.	
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		
	eta 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 $eta=\mu$.	
Un-selected is correct		
~	1 / 1 point	
	raining a neural network with Batch Norm, at test time, to evaluate the neural network on a new le you should:	
	Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.	