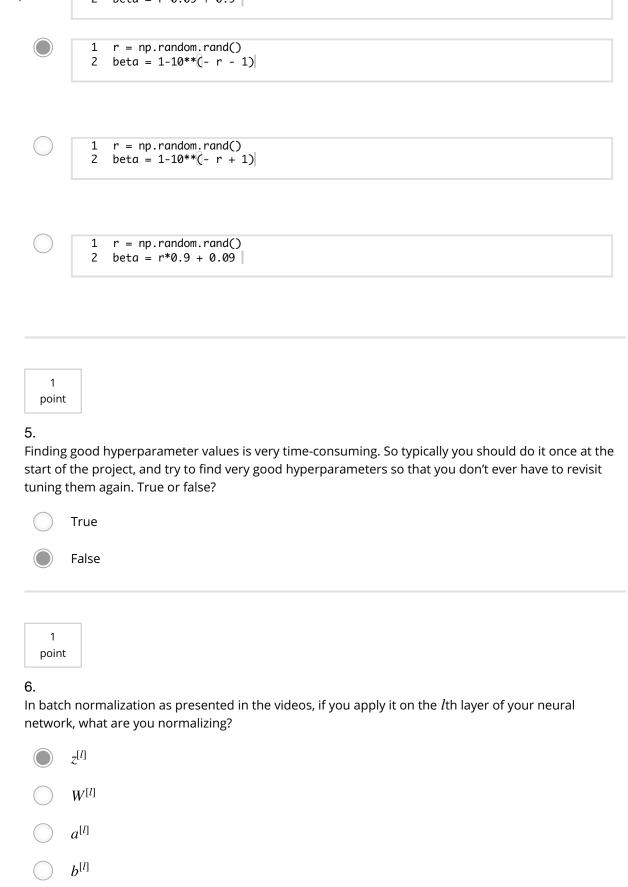
Hyperparameter tuning, Batch Normalization, Programming Frameworks

Quiz, 10 questions

| 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? True False |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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? True False |
| 1 point 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: Whether you use batch or mini-batch optimization The presence of local minima (and saddle points) in your neural network The amount of computational power you can access The number of hyperparameters you have to tune |
| 1 point |

4.



| 7. | |
|-----------------------------|------------------------------------------|
| Hypenparametertuning, Batel | n Natrmalization paragramming Frameworks |

Quiz, 10 questions

| | To avoid division by zero |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | In case μ is too small |
| | To have a more accurate normalization |
| | To speed up convergence |
| | |
| 1 point | |
| 8. | |
| Which | of the following statements about γ and eta in Batch Norm are true? |
| | They set the mean and variance of the linear variable $z^{\mathrm{I}}l$ of a given layer. |
| | 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. |
| | They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent. |
| | eta and γ are hyperparameters of the algorithm, which we tune via random sampling. |
| | The optimal values are $\gamma=\sqrt{\sigma^2+\varepsilon}$, and $\beta=\mu$. |
| | |
| 1 point | t |
| 9. | |
| | raining a neural network with Batch Norm, at test time, to evaluate the neural network on a new le you should: |
| | 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. |
| | Perform the needed normalizations, use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training. |
| | Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations. |
| | Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized. |

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Quiz, 10 quelions
 Which of these statements about deep learning programming frameworks are true? (Check all that apply)

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

 Upgrade to submit

