✓ Congratulations! You passed! Grade received 90% To pass 80% or higher	Go to next item
Hyperparameter tuning, Batch Normalization, Program  Latest Submission Grade 90%	ming Frameworks
1. Which of the following are true about hyperparameter search?	1/1 point
<ul> <li>When using random values for the hyperparameters they must be always uniformly distributed.</li> <li>When sampling from a grid, the number of values for each hyperparameter is larger than</li> </ul>	
when using random values.  Choosing values in a grid for the hyperparameters is better when the number of	
hyperparameters to tune is high since it provides a more ordered way to search.  Choosing random values for the hyperparameters is convenient since we might not know in advance which hyperparameters are more important for the problem at hand.	
∠ <sup>¬</sup> Expand	
Correct  Correct. Different problems might be more sensitive to different hyperparameters.	
2. If it is only possible to tune two parameters from the following due to limited computational result would you choose?	sources. Which two 1/1 point
$leftilde{\square}$ $\alpha$	
Correct Correct. This might be the hyperparameter that most impacts the results of a model.	
$\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $	
$\checkmark$ The $_{eta}$ parameter of the momentum in gradient descent.	
<ul> <li>Correct</li> <li>Correct. This hyperparameter can increase the speed of convergence of the training, thus is worth tuning.</li> </ul>	
∠ <sup>7</sup> Expand	
Correct Great you got all the right answers	
Great, you got all the right answers.	
3. Even if enough computational power is available for hyperparameter tuning, it is always better model ("Panda" strategy), since this will result in a more custom model. True/False?	to babysit one 1/1 point
O True	
₹ <b>5</b>	
✓ Expand ✓ Correct	
Correct. Although it is possible to create good models using the "Panda" strategy, obtaining is more likely using a "caviar" strategy due to the number of tests and the nature of the deprocess of ideas, code, and experiment.	
<b>4.</b> Knowing that the hyperparameter $\alpha$ should be in the range of $0.001$ and $1.0$ . Which of the following recommended way to sample a value for $\alpha$ ?	lowing is the 1/1 point
r = -3*np.random.rand() alpha = 10**r	
r = np.random.rand() alpha = 0.001 + r*0.999  r = 4*np.random.rand()	
alpha = 10**r  r = -5*np.random.rand()  alpha = 10**r	
∠ <sup>⊅</sup> Expand	
$\bigcirc$ Correct  Yes. This gives a random number between $0.001 = 10^{-3}$ and $10^{0}$ .	
res. This gives a random number between 0.001 = 10 and 10.	
5. Finding new values for the hyperparameters, once we have found good ones for a model, should new hardware or computational power is acquired. True/False?	ld only be done if 1/1 point
False	
O True	
√ <sup>7</sup> Expand	
<ul> <li>✓ Correct</li> <li>Correct. As the data changes for the model, it might be beneficial to tune some of the hyp</li> </ul>	perparameters
again.	oct parameters
6. When using batch normalization it is OK to drop the parameter $W^{[l]}$ from the forward propagat	tion since it will be 1/1 point
subtracted out when we compute $ ilde z_{ m normalize}^{[l]}=eta^{[l]}\hat z^{[l]}+\gamma^{[l]}$ . True/False? $igcap$	
False	
∠ Expand  ✓ Correct	
Correct. The parameter $W^{[I]}$ doesn't get subtracted during the batch normalization procegets re-scaled.	ess, although it
(2) (8)	
7. In the normalization formula $z_{norm}^{(i)}=\frac{z^{(i)}-\mu}{\sqrt{\sigma^2+\varepsilon}}$ , why do we use epsilon? $\bigcirc \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ $	1/1 point
To avoid division by zero	
To speed up convergence  To have a more accurate normalization	
∠ <sup>¬</sup> Expand	
⊘ Correct	
8. Which of the following are true about batch normalization?	1/1 point
$\ \ \ \ _{eta^{[l]}}$ and $_{\gamma^{[l]}}$ are hyperparameters that must be tuned by random sampling in a logarithmic	± / ± point
scale.	
$\sqrt{\sigma^2}$ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the mean and variance of $\widetilde{z}^{[l]}$ .	
Correct $ \hbox{Correct. When applying the linear transformation } \tilde{z}^{(l)} = \beta^{[l]} z^{(l)}_{norm} + \gamma^{[l]} \hbox{ we set the } $ mean and variance of $z^{[l]}$ .	
mean and variance of $_{\widetilde{z}}[l]$ . $ \qquad \qquad$	
"learned" or trained.  Correct	
Correct. Batch normalization uses two parameters $_{eta}$ and $_{\gamma}$ to compute $\tilde{z}^{(i)}=\beta z_{norm}^{(i)}+\gamma\cdot$	
∠ Expand  ✓ Correct	
Great, you got all the right answers.	
9. A neural network is trained with Batch Norm. At test time, to evaluate the neural network we turn Norm to avoid random predictions from the network. True/False?	rn off the Batch 0 / 1 point
Norm to avoid random predictions from the network. True/False?  False	
True	
Expand  Note: The second control of the seco	
$igotimes$ Incorrect Incorrect. During the test, the parameters $\mu$ and $\sigma^2$ are estimated using an exponentially value average across mini-batches used during training.	weighted
10. If a project is open-source, it is a guarantee that it will remain open source in the long run and we modified to benefit only one company. True/False?	will never be 1/1 point
○ True ♠ False	
∠ <sup>¬</sup> Expand	
<ul> <li>✓ Correct</li> <li>Correct. To ensure that a project will remain open source in the long run it must have a go</li> </ul>	ood governance
body too.	