 When using random values for the hyperparameters they must be always uniformly distributed. 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. 	
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.	
When sampling from a grid, the number of values for each hyperparameter is larger than when using random values.	
∠ [¬] 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 resources. Which two would you choose?	1/1 point
The $_{oldsymbol{eta}}$ parameter of the momentum in gradient descent.	
✓ Correct Correct. This hyperparameter can increase the speed of convergence of the training, thus is worth tuning.	
$\ \ \ \ \epsilon$ in Adam.	
$ec{}\!$	
✓ Correct Correct. This might be the hyperparameter that most impacts the results of a model.	
$oxed{\ }$ eta_1' eta_2 in Adam.	
∠ [™] Expand	
Correct Great, you got all the right answers.	
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: 1/1 point
Whether you use batch or mini-batch optimization	
The number of hyperparameters you have to tune	
The amount of computational power you can access	
The presence of local minima (and saddle points) in your neural network	
∠ [™] Expand	
Knowing that the hyperparameter $lpha$ should be in the range of 0.001 and 1.0 . Which of the following is the recommended way to sample a value for $lpha$?	0 / 1 point
r = -5*np.random.rand() alpha = 10**r	
r = np.random.rand() $alpha = 0.001 + r*0.999$	
∠ [™] Expand	
$igotimes$ Incorrect No. This will generate a random value between 10^{-5} and 10^{0} .	
No. This will generate a random value between 10 and 10.	
Finding new values for the hyperparameters, once we have found good ones for a model, should only be done if new hardware or computational power is	acquired. 1/1 point
True/False? True	
False	
∠ [™] Expand	
⊘ Correct	
Correct. As the data changes for the model, it might be beneficial to tune some of the hyperparameters again.	
6. When using batch normalization it is OK to drop the parameter $W^{[l]}$ from the forward propagation since it will be subtracted out when we compute	1/1 point
$ ilde{z}^{[l]} = \gamma z_{ ext{normalize}}^{[l]} + eta^{[l]}$. True/False?	
○ True	
False	
∠ [™] Expand	
Correct. The parameter $W^{[l]}$ doesn't get subtracted during the batch normalization process, although it gets re-scaled.	
. In the normalization formula $z^{(i)} = z^{(i)} - \mu$, why do we use ancilar?	1/1 point
' In the normalization formula $z_{norm}^{(i)}=rac{z^{(i)}-\mu}{\sqrt{\sigma^2+arepsilon}}$, why do we use epsilon?	1/1 point
	, ,
O In case μ is too small To have a more accurate normalization	
In case μ is too small To have a more accurate normalization To speed up convergence	
To have a more accurate normalization	
To have a more accurate normalization To speed up convergence	
To have a more accurate normalization To speed up convergence To avoid division by zero	
To have a more accurate normalization To speed up convergence To avoid division by zero Expand	
To have a more accurate normalization To speed up convergence To avoid division by zero	
To have a more accurate normalization To speed up convergence To avoid division by zero Expand ✓ Correct	
To have a more accurate normalization To speed up convergence To avoid division by zero ✓ Expand ✓ Correct Which of the following is true about batch normalization?	
 To have a more accurate normalization To speed up convergence To avoid division by zero Expand ✓ Correct Which of the following is true about batch normalization? The parameters γ μ and β μ set the variance and mean of z̄ μ.	
To have a more accurate normalization To speed up convergence To avoid division by zero To avoid division by zero Correct Which of the following is true about batch normalization? The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ set the variance and mean of $z_{[i]}$. The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.	
 To have a more accurate normalization To speed up convergence To avoid division by zero Expand ✓ Correct Which of the following is true about batch normalization? The parameters γ μ and β μ set the variance and mean of z μ.	
To have a more accurate normalization To speed up convergence To avoid division by zero To avoid division by zero To avoid division by zero Correct Which of the following is true about batch normalization? The parameters $\gamma_{[l]}$ and $\beta_{[l]}$ set the variance and mean of $z_{[l]}$. The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$. $z_{norm}^{(l)} = \frac{z^{(l)} - \mu}{\sqrt{\sigma^2}}$.	
To have a more accurate normalization To speed up convergence To avoid division by zero To avoid division by zero To avoid division by zero Correct Which of the following is true about batch normalization? The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ set the variance and mean of $z_{[i]}$. The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$. $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.	
To have a more accurate normalization ○ To speed up convergence ⑤ To avoid division by zero ② Correct 3. Which of the following is true about batch normalization? ⑥ The parameters $_{\gamma}$ [$_{ij}$] and $_{\beta}$ [$_{ij}$] set the variance and mean of $_{z}$ [$_{ij}$]: ○ The optimal values to use for $_{\gamma}$ and $_{\beta}$ are $_{\gamma} = \sqrt{\sigma^2 + \epsilon}$ and $_{\beta} = \mu$. ○ $_{z_{torm}} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.	
To have a more accurate normalization To speed up convergence To avoid division by zero To avoid division by zero To avoid division by zero Correct Which of the following is true about batch normalization? The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ set the variance and mean of $z_{[i]}$. The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$. $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.	
To have a more accurate normalization To speed up convergence To avoid division by zero To avoid division by zero Correct Which of the following is true about batch normalization? The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ set the variance and mean of $z_{[i]}$. The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$. $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.	
To have a more accurate normalization To speed up convergence To avoid division by zero The following is true about batch normalization? The parameters $\gamma^{[i]}$ and $\beta^{[i]}$ set the variance and mean of $\gamma^{[i]}$ and be learned only using plain gradient descent.	
To have a more accurate normalization ○ To speed up convergence ③ To avoid division by zero ✓ Correct Which of the following is true about batch normalization? ④ The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ set the variance and mean of $\frac{1}{2}\beta_{[i]}$. ○ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$: ○ $\frac{\epsilon_{00}}{2\cos m} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$. ○ The parameters $\gamma_{[i]}$ and $\beta_{[i]}$ can be learned only using plain gradient descent.	
○ To have a more accurate normalization ○ To speed up convergence ② To avoid division by zero ② Correct Which of the following is true about batch normalization? ③ The parameters $_{\gamma \beta}$ and $_{\beta \beta }$ set the variance and mean of $_{\beta \beta}$. ○ The optimal values to use for $_{\gamma}$ and $_{\beta}$ are $_{\gamma} = \sqrt{\sigma^3 + \epsilon}$ and $_{\beta} = \mu$. ○ $_{\alpha}^{00} = \frac{z^{(0)} - \mu}{\sqrt{\sigma^3}}$. ○ The parameters $_{\gamma \beta}$ and $_{\beta \beta }$ can be learned only using plain gradient descent. ② Correct Correct. When applying the linear transformation $_{\Sigma}^{(f)} = \beta^{[h]} z_{norm}^{(f)} + \gamma^{[f]}$ we set the variance and mean of $_{\Sigma}^{(f)}$.	1/1 point
To have a more accurate normalization To speed up convergence To avoid division by zero ■ To avoid division by zero ■ To avoid division by zero ■ The parameters $\gamma_{[i]}$ and $\gamma_{[i]}$ set the variance and mean of $\gamma_{[i]}$. The optimal values to use for $\gamma_{[i]}$ and $\gamma_{[i]}$ set the variance and mean of $\gamma_{[i]}$. The optimal values to use for $\gamma_{[i]}$ and $\gamma_{[i]}$ set $\gamma_{[i]}$ and $\gamma_{[i]}$ can be learned only using plain gradient descent. ■ The parameters $\gamma_{[i]}$ and $\gamma_{[i]}$ can be learned only using plain gradient descent. ■ Correct Correct. When applying the linear transformation $\tilde{x}^{[i]} = \beta^{[i]} z_{norm}^{[i]} + \gamma^{[i]}$ we set the variance and mean of $\tilde{x}^{[i]}$. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should: If you implemented Batch Norm on mini-batches of (asy) 256 examples, then to evaluate on one test example, duplicate that	1/1 point
To have a more accurate normalization ○ To speed up convergence ⑥ To avoid division by zero ○ Correct • Which of the following is true about batch normalization? ⑥ The parameters $\gamma_{[l]}$ and $\beta_{[l]}$ set the variance and mean of $\frac{1}{2}$ [θ_{l}]. ○ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^3 + \epsilon}$ and $\beta = \mu$. ○ $\frac{1}{2}$ [$\frac{1}{2}$]. ○ The parameters $\gamma_{[l]}$ [$\frac{1}{2}$] and $\beta_{[l]}$ [$\frac{1}{2}$] can be learned only using plain gradient descent. ○ Correct ○ Correct ○ Correct When applying the linear transformation $\frac{1}{2}$ ($\frac{1}{2}$) = $\frac{1}{2}$ [$\frac{1}{2}$], we set the variance and mean of $\frac{1}{2}$ [$\frac{1}{2}$]. • 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
To have a more accurate normalization To speed up convergence To avoid division by zero The parameters γ _R and γ _R set the variance and mean of γ _R to the optimal values to use for γ and ρ are γ = √σ³ + ε and ρ = μ' γ ^R The optimal values to use for γ and ρ are γ = √σ³ + ε and ρ = μ' γ ^R The parameters γ _R and γ _R can be learned only using plain gradient descent. The parameters γ _R and γ _R can be learned only using plain gradient descent. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example 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. Use the most recent mini-batch's value of μ and σ _R to perform the needed normalizations. Skip the step where you normalize using μ and σ _R a lange test example cannot be normalized.	1/1 point
To have a more accurate normalization To speed up convergence To avoid division by zero The parameters $_{\gamma i }$ and $_{\beta i }$ set the variance and mean of $_{\tilde{z}}$ iii. The populariest values to use for $_{\gamma}$ and $_{\beta}$ set $_{\gamma}$ $_{\gamma}$ $_{\gamma}$ and $_{\beta}$ set $_{\gamma}$ $_$	1/1 point
 To have a more accurate normalization To speed up convergence To avoid division by zero Which of the following is true about batch normalization? The parameters _γ g and _{gift} set the variance and mean of ½0. The optimal values to use for _γ and _g are _γ = √σ² + ε and g = μ' 20 and γ = √σ². The parameters _γ g and _{gift} can be learned only using plain gradient descent. The parameters _γ g and _{gift} can be learned only using plain gradient descent. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should: If you implemented Batch Norm on mini-batches of (pay) 256 examples, then to evaluate on one test example 256 times so that you've evolving with a min-batch the same size as during training. Use the most recent mini-batch's value of μ and σ to perform the needed normalizations. Stip the step where you normalize using μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. 	1/1 point
 To have a more accurate normalization To speed up convergence To avoid division by zero Which of the following is true about batch normalization? The parameters _γ g and _{gift} set the variance and mean of ½0. The optimal values to use for _γ and _g are _γ = √σ² + ε and g = μ' 20 and γ = √σ². The parameters _γ g and _{gift} can be learned only using plain gradient descent. The parameters _γ g and _{gift} can be learned only using plain gradient descent. After training a neural network with Batch Norm, at test time, to evaluate the neural network on a new example you should: If you implemented Batch Norm on mini-batches of (pay) 256 examples, then to evaluate on one test example 256 times so that you've evolving with a min-batch the same size as during training. Use the most recent mini-batch's value of μ and σ to perform the needed normalizations. Stip the step where you normalize using μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. Perform the needed normalizations, use μ and σ z since a single test example cannot be normalized. 	1/1 point
To have a more accurate normalization To speed up convergence To avoid division by zero The parameters γ _R and ρ _R set the variance and mean of gR- The optimal values to use for γ and ρ are γ = √√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√	1/1 point
 To have a more accurate connatization To speed up convergence To avoid division by zero Which of the following is true about batch normalization? The parameters y_k and y_B set the variance and mean of y_B The optimal values to use for y and y_B ≈ 0 y − √√3±√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√	1/1 point
 To have a more accurate normalization To upsed up convergence To avoid division by zero Which of the following is true about batch normalization? The parameters y_k and y_k set the variance and mean of zip. The coolinal values to use for y and y after y = √√x 1 x and y = µ. z̄ z̄ z	1/1 point
 To have a more accurate normalization To speed up crowagenze To avoid division by zero The parameters y_R and y_R set the variance and mean of z_R to the cotinn values to use for y_R and y_R set y_R = √x² + e and y = µ' The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using plan gradient descent. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The parameters y_R and y_R can be learned only using y_R y^R we set the variance and mean of z̄^R. The param	1/1 point
To have a more accurate normalization	1/1 point
To a proof a counter accounts normalization	1/1 point
To have a more accurate reconstruction To sever dishelden by zero The parameters γ _{ij} and γ _{ij} set the variance and mean of χ _{ij} . The contined valves to use for γ _i and γ _{ij} are γ _{ij} ∨ √√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√√	1/1 point
To have a more accurate reconstitution	1/1 point
The have a more according control process The process of the control of the	1/1 point
So have a more securitie recomalization	1/1 point

<