

1. Which of the following are true about hyperparameter search?

1 / 1 point

- ☒ 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.
- ☐ When using random values for the hyperparameters they must be always uniformly distributed.
- ☐ When sampling from a grid, the number of values for each hyperparameter is larger than 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.

Expand

Correct

Correct. Different problems might be more sensitive to different hyperparameters.

Activate Windows

2. If it is only possible to tune two parameters from the following due to limited computational resources. Which two would you choose?

0 / 1 point

☒ β_1, β_2 in Adam.

! This should not be selected
Incorrect. This hyperparameter has little impact and it is usually better to use the default values 0.9, 0.999.

☐ The β parameter of the momentum in gradient descent.

☐ ϵ in Adam.

☒ α

Correct

Correct. This might be the hyperparameter that most impacts the results of a model.

Expand

Incorrect

You didn't select all the correct answers.

Activate Windows
Go to Settings to activate Windows.

3. Using the "Panda" strategy, it is possible to create several models. True/False?

0 / 1 point

☐ True

☒ False

Expand

Incorrect

Incorrect. Following the "Panda" analogy, it is possible to babysit a model until a certain point and then start again to produce a different one.

Activate Windows
Go to Settings to activate Windows.

4. Knowing that the hyperparameter α should be in the range of 0.00001 and 1.0, which of the following is the recommended way to sample a value for α ?

0 / 1 point

- ☐ $r = \text{np.random.rand}()$
 $\alpha = 0.00001 + r * 0.99999$
- ☒ $r = -4 * \text{np.random.rand}()$
 $\alpha = 10^{**}r$
- ☐ $r = \text{np.random.rand}()$
 $\alpha = 10^{**}r$
- ☐ $r = -5 * \text{np.random.rand}()$
 $\alpha = 10^{**}r$

[Expand](#)

✗ Incorrect

No. This gives a random number between 10^{-4} and 10^0 .

Activate Windows
Go to Settings to activate Windows.

5. 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. True/False?

1 / 1 point

- ☐ True
- ☒ False

[Expand](#)

✓ Correct

Correct. As the data changes for the model, it might be beneficial to tune some of the hyperparameters again.

Activate Windows
Go to Settings to activate Windows.

Since 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 $b^{[l]}$ from the forward propagation since it will be subtracted out when we compute $z^{[l]} = \gamma z_{\text{normalize}}^{[l]} + \beta^{[l]}$. True/False?

0 / 1 point

- ☐ True
- ☒ False

[Expand](#)

✗ Incorrect

Incorrect. Since in the normalization process the values of $z^{[l]}$ are re-centered at the origin, it is irrelevant to add the $b^{[l]}$ parameter.

Activate Windows
Go to Settings to activate Windows.

7. In the normalization formula $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$, why do we use epsilon?

1 / 1 point

- ☐ In case μ is too small
- ☐ To have a more accurate normalization
- ☒ To avoid division by zero
- ☐ To speed up convergence

Expand

Correct

Activate Windows
Go to Settings to activate Windows.

8. Which of the following is true about batch normalization?

1 / 1 point

- ☐ $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- ☐ The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.
- ☒ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\hat{x}^{[l]}$.
- ☐ The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.

Expand

Correct

Correct. When applying the linear transformation $\hat{z}^{(l)} = \beta^{[l]} z_{norm}^{(l)} + \gamma^{[l]}$ we set the variance and mean of $\hat{z}^{[l]}$.

Activate Windows
Go to Settings to activate Windows.

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

- ☐ 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^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.
- ☐ Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.

Expand

Correct

Activate Windows

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

1 / 1 point

- ☒ Even if a project is currently open source, good governance of the project helps ensure that it remains open even in the long term, rather than become closed or modified to benefit only one company.

✓ Correct

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

- ☒ 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.

✓ Correct

[Expand](#)

✓ Correct

Great, you got all the right answers.

Activate Windows
Go to Settings to activate Windows.

2:58 AM