

✓ Congratulations! You passed!

Go to next item

Grade received 100% Latest Submission Grade 100% To pass 80% or higher

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?

1 / 1 point

☒ False

☐ True

↗ Expand

✓ Correct

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

☐ ϵ in Adam.

☒ The β parameter of the momentum in gradient descent.

✓ Correct

Correct. This hyperparameter can increase the speed of convergence of the training, thus is worth tuning.

☒ α

✓ Correct

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

☐ β_1, β_2 in Adam.

↗ Expand

✓ Correct

Great, you got all the right answers.

3. Even if enough computational power is available for hyperparameter tuning, it is always better to babysit one model ("Panda" strategy), since this will result in a more custom model. True/False?

1 / 1 point

☒ False

☐ True

↗ Expand

✓ Correct

Correct. Although it is possible to create good models using the "Panda" strategy, obtaining better results is more likely using a "caviar" strategy due to the number of tests and the nature of the deep learning process of ideas, code, and experiment.

4. Knowing that the hyperparameter α should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for α ?

1 / 1 point

- ☐ $r = -5 * \text{np.random.rand()}$
 $\alpha = 10^{**}r$
- ☐ $r = \text{np.random.rand()}$
 $\alpha = 0.001 + r * 0.999$
- ☒ $r = -3 * \text{np.random.rand()}$
 $\alpha = 10^{**}r$
- ☐ $r = 4 * \text{np.random.rand()}$
 $\alpha = 10^{**}r$

[Expand](#)

✓ Correct

Yes. This gives a random number between $0.001 = 10^{-3}$ and 10^0 .

5. Finding good hyperparameter values is very time-consuming. So typically you should do it once at the start of the project, and try to find very good hyperparameters so that you don't ever have to tune them again. True or false?

1 / 1 point

- ☐ True
- ☒ False

[Expand](#)

✓ Correct

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 $\tilde{z}^{[l]} = \gamma z_{\text{normalize}}^{[l]} + \beta^{[l]}$. True/False?

1 / 1 point

- ☒ True
- ☐ False

[Expand](#)

✓ Correct

Correct. 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.

7. When using normalization:

1 / 1 point

$$z_{\text{norm}}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

In case σ is too small, the normalization of $z^{(i)}$ may fail since division by 0 may be produced due to rounding errors. True/False?

- ☐ True
- ☒ False

Expand

Correct

Correct. The normalization formula uses a smoothing parameter ϵ so in $z_{\text{norm}}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$ use of the ϵ parameter prevents that the denominator be 0.

8. Which of the following statements about γ and β in Batch Norm are true?

1 / 1 point

- ☐ The optimal values are $\gamma = \sqrt{\sigma^2 + \epsilon}$, and $\beta = \mu$.
- ☒ They set the mean and variance of the linear variable $z^{[l]}$ of a given layer.

Correct

- ☐ There is one global value of

$$\gamma \in \mathbb{R}$$

$\gamma \in \mathbb{R}$ and one global value of

$$\beta \in \mathbb{R}$$

$\beta \in \mathbb{R}$ for each layer, and these apply to all the hidden units in that layer.

- ☐ β

β and

Expand

Correct

Great, you got all the right answers.

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

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

Expand

Correct

10. Which of the following are some recommended criteria to choose a deep learning framework?

1 / 1 point

- ☐ It must be implemented in C to be faster.
- ☒ Running speed.
- ☐ It must use Python as the primary language.
- ☐ It must run exclusively on cloud services, to ensure its robustness.

Expand

✓ **Correct**

Correct. The running speed is a major factor, especially when working with large datasets.