

✓ **Congratulations! You passed!**

Grade received 90% Latest Submission Grade 90% To pass 80% or higher

Retake the assignment in  
23h 49m

Go to next  
item

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

1 / 1 point

- ☐ 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.
- ☒ 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.

↗ Expand



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

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?

1 / 1 point

- ☐ True
- ☒ False

↗ Expand



Yes. We've seen in the lecture that some hyperparameters, such as the learning rate, are more critical than others.

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

1 / 1 point

- ☒ True
- ☐ False

↗ Expand



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

4. If you think  $\beta$  (hyperparameter for momentum) is between 0.9 and 0.99, which of the following is the recommended way to sample a value for beta?

1 / 1 point

- ☐  $r = \text{np.random.rand}()$   
 $\text{beta} = r * 0.09 + 0.9$
- ☐  $r = \text{np.random.rand}()$   
 $\text{beta} = 1 - 10^{**}(-r + 1)$

- ☒ `r = np.random.rand()`  
`beta = 1-10**(- r - 1)`
- ☐ `r = np.random.rand()`  
`beta = r*0.9 + 0.09`

[Expand](#)

✓ Correct

5. Once good values of hyperparameters have been found, those values should be changed if new data is added or a change in computational power occurs. True/False?

1 / 1 point

- ☐ False
- ☒ True

[Expand](#)

✓ Correct

Correct. The choice of some hyperparameters such as the batch size depends on conditions such as hardware and quantity of data.

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  $\hat{z}_{\text{normalize}}^{[l]} = \beta^{[l]} \hat{z}^{[l]} + \gamma^{[l]}$ . 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:

0 / 1 point

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

In case  $\sigma$  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](#)

✗ Incorrect

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

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

1 / 1 point

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

Expand

✓ Correct

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

9. A neural network is trained with Batch Norm. At test time, to evaluate the neural network we turn off the Batch Norm to avoid random predictions from the network. True/False?

1 / 1 point

- ☐ True
- ☒ False

Expand

✓ Correct

Correct. During the test, the parameters  $\mu$  and  $\sigma^2$  are estimated using an exponentially weighted average across mini-batches used during training.

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

- ☒ 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

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

Expand

✓ Correct

Great, you got all the right answers.