

**Your grade: 100%****Next item →**

Your latest: 100% • Your highest: 100% • To pass you need at least 80%. We keep your highest score.

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

**Correct**

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

- False
- True

**Correct**

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

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

- True
- False

**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  $\alpha$  should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for  $\alpha$ ? 1 / 1 point

- 
- ```
r = 4*np.random.rand()
```

  

```
alpha = 10**r
```
- 
- ```
r = np.random.rand()
```

  

```
alpha = 0.001 + r*0.999
```
- 
- ```
r = -5*np.random.rand()
```

  

```
alpha = 10**r
```
- 
- ```
r = -3*np.random.rand()
```

  

```
alpha = 10**r
```

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

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

**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  $W^{[l]}$  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

**Correct**Correct. The parameter  $W^{[l]}$  doesn't get subtracted during the batch normalization process, although it gets re-scaled.

7. In the normalization formula  $z_{\text{norm}}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$ , why do we use epsilon? 1 / 1 point

- To have a more accurate normalization
- In case  $\mu$  is too small
- To speed up convergence
- To avoid division by zero

**Correct**

8. Which of the following is true about batch normalization? 1 / 1 point

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

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

- False
- True

**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. If a project is open-source, it is a guarantee that it will remain open source in the long run and will never be modified to benefit only one company. True/False? 1 / 1 point

- True
- False

**Correct**

Correct. To ensure that a project will remain open source in the long run it must have a good governance body too.