

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

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.
- 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 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.

[Expand](#)

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?

0 / 1 point

- False
- True

[Expand](#)

Incorrect

No. We've seen in the lecture that some hyperparameters, such as the learning rate, are more critical than others.

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

- The amount of computational power you can access
- Whether you use batch or mini-batch optimization
- The number of hyperparameters you have to tune
- The presence of local minima (and saddle points) in your neural network

[Expand](#)

Correct

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 = 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`

 Expand

 Correct

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

 Correct

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

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

- False
- True

 Expand

 Correct

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

6. In batch normalization as presented in the videos, if you apply it on the l th layer of your neural network, what are you normalizing?

1 / 1 point

- $z^{[l]}$
- $W^{[l]}$
- $a^{[l]}$
- $b^{[l]}$

 Expand

 Correct

7. When using normalization:

1 / 1 point

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

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?

- False
- True

 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 is true about batch normalization?

1 / 1 point

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

 Expand

 Correct

Correct. When applying the linear transformation $\tilde{z}^{(l)} = \beta^{[l]} z_{\text{norm}}^{(l)} + \gamma^{[l]}$ 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 on a new example you should perform the normalization using μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training. True/false?

1 / 1 point

- True
- False

 Expand

 Correct

Correct. This is a good practice to estimate the μ and σ^2 to use since at test time we might not be predicting over a batch of the same size, or it might even be a single example, thus using the μ and σ^2 of a single sample doesn't make sense.

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

1 / 1 point

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

 Expand

 Correct

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