

✓ **Congratulations! You passed!**

Grade  
received **80%**

Latest Submission  
Grade 80%

To pass 80% or  
higher

Go to next item

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

1 / 1 point

- ☐ 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.
- ☐ 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. In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply.

0 / 1 point

- ☐  $\epsilon$  in Adam.
- ☒ mini-batch size

✓ **Correct**

Correct. This can have a great impact on the results of the cost function, thus it is worth tuning it.

- ☐ The  $\beta$  parameter of the momentum in gradient descent.
- ☒  $\alpha$

✓ **Correct**

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

- ☐  $\beta_1, \beta_2$  in Adam.

↗ Expand

✗ **Incorrect**

You didn't select all the correct answers

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  $\alpha$  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

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

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

- ☒ False
- ☐ True

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. Which of the following are true about batch normalization?

0 / 1 point

- ☐ There is a global value of  $\gamma$  and  $\beta$  that is used for all the hidden layers where batch normalization is used.
- ☐ The parameters  $\beta$  and  $\gamma$  of batch normalization can't be trained using Adam or RMS prop.
- ☒ The parameter  $\epsilon$  in the batch normalization formula is used to accelerate the convergence of the model.
- ☐ One intuition behind why batch normalization works is that it helps reduce the internal covariance.

Expand

✗ Incorrect

No. The  $\epsilon$  parameter is used to avoid division by 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 parameters  $\gamma^{[l]}$  and  $\beta^{[l]}$  can be learned only using plain gradient descent.
- ☐ The optimal values to use for  $\gamma$  and  $\beta$  are  $\gamma = \sqrt{\sigma^2 + \epsilon}$  and  $\beta = \mu$ .
- ☐  $z_{norm}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$ .

 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. 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

- ☐ Use the most recent mini-batch's value of  $\mu$  and  $\sigma^2$  to perform the needed normalizations.
- ☐ Skip the step where you normalize using  $\mu$  and  $\sigma^2$  since a single test example cannot be normalized.
- ☒ Perform the needed normalizations, use  $\mu$  and  $\sigma^2$  estimated using an exponentially weighted average across mini-batches seen during training.
- ☐ 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.

 Expand

 **Correct**

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

1 / 1 point

- ☒ Running speed.
- ☐ It must be implemented in C to be faster.
- ☐ 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.