Quiz: Hyperparameter tuning, Batch Normalization, Programming Frameworks

✅ Congratulations! You passed!

Grade	Latest Submission	To pass 80% or	**Go to next item**
received 90%	Grade 90%	higher	

---

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 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.

   ⤢ Expand

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

---

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**

   ☐ β_1, β_2 in Adam.

   ☐ 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.

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

⤢ 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*np.random.rand()
alpha = 10**r

○ r = np.random.rand()
alpha = 0.001 + r*0.999

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

○ r = 4*np.random.rand()
alpha = 10**r

⤢ Expand

⊘ **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?

0 / 1 point

○ False

○ True

[Expand]

⊗ **Incorrect**
Incorrect. The choice of some hyperparameters such as the batch size, depends on conditions such as hardware and quantity of data.

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

1 / 1 point

○ $W^{[l]}$

○ $a^{[l]}$

◉ $z^{[l]}$

○ $b^{[l]}$

[Expand]

✓ **Correct**

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

○ To have a more accurate normalization

○ In case μ is too small

○ To speed up convergence

◉ To avoid division by zero

⤢ Expand

⊘ Correct

8. Which of the following statements about γ and β in Batch Norm are true?   1 / 1 point

☑ They can be learned using Adam, Gradient descent with momentum, or RMSprop, not just with gradient descent.

✓ Correct

☐ The optimal values are $\gamma = \sqrt{\sigma^2 + \varepsilon}$, and $\beta = \mu$.

☑ They set the variance and mean of the linear variable $z^{[l]}$ of a given layer.

✓ Correct

☐ β and γ are hyperparameters of the algorithm, which we tune via random sampling.

☐ There is one global value of $\gamma \in \mathfrak{R}$ and one global value of $\beta \in \mathfrak{R}$ for each layer, and these apply to all the hidden units in that layer.

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

○ 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.

○ Use the most recent mini-batch's value of μ and σ^2 to perform the needed normalizations.

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

◉ Perform the needed normalizations, use μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training.

⤢ Expand

✓ Correct

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

◉ False

○ True

⤢ Expand

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