Congratulations! You passed!

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

Go to next item

1. If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False? **1 / 1 point**

- ○ True
- ◉ False

Expand

Correct

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**

- ☐ 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.

- ☐ β_1, β_2 in Adam.
- ☑ α

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

Expand

Correct
Great, you got all the right 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
- ○ The number of hyperparameters you have to tune
- ○ Whether you use batch or mini-batch optimization
- ○ The presence of local minima (and saddle points) in your neural network

Expand

Correct

4. If you think β (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 = np.random.rand()
beta = 1-10**(- r - 1)
- ○ r = np.random.rand()
beta = r*0.9 + 0.09
- ○ r = np.random.rand()
beta = r*0.09 + 0.9
- ○ r = np.random.rand()
beta = 1-10**(- r + 1)

Expand

Correct

5. Finding new values for the hyperparameters, once we have found good ones for a model, should only be done if **1 / 1 point**

new hardware or computational power is acquired. True/False?

- ○ True
- ◉ False

Expand

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

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 $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False? **0 / 1 point**

- ○ True
- ◉ False

Expand

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

- ○ To speed up convergence
- ◉ To avoid division by zero
- ○ In case μ is too small
- ○ To have a more accurate normalization

Expand

✓ **Correct**

8. Which of the following are true about batch normalization? **0 / 1 point**

- ☑ When using batch normalization we introduce two new parameters $\gamma^{[l]}$, $\beta^{[l]}$ that must be "learned" or trained.

  ✓ **Correct**
  Correct. Batch normalization uses two parameters β and γ to compute $\tilde{z}^{(i)} = \beta z^{(i)}_{norm} + \gamma$.

- ☑ $\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.

  ! **This should not be selected**
  Incorrect. These are parameters that must be learned. They can be learned with gradient descent, gradient descent with momentum, RMSprop, and Adam, like all the other parameters.

- ☐ $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.

- ☐ The parameters $$\gamma^{[l]}$$ and $$\beta^{[l]}$$ set the variance and mean of $$\widetilde{z}^{[l]}$$.

Expand

✗ **Incorrect**
You didn't select all the correct answers

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**

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