Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

1. With a relatively small set of hyperparameters, it is OK to use a grid search. True/False? **1 / 1 point**

- (•) True
- ( ) False

Expand

Correct
Correct. When the set of hyperparameters is small like a range for $n_l = 1, 2, 3$ grid search works fine.

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**

- [x] α

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

- [ ] ϵ in Adam.
- [ ] β_1, β_2 in Adam.
- [x] 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.

Expand

Correct
Great, you got all the right answers.

3. Using the "Panda" strategy, it is possible to create several models. True/False? **1 / 1 point**

- ( ) False
- (•) True

Expand

Correct
Correct. Following the "Panda" analogy, it is possible to babysit a model until a certain point and then start again to produce a different one

Correct. Following the "Panda" analogy, it is possible to babysit a model until a certain point and then start again to produce a different one.

4. Knowing that the hyperparameter α should be in the range of 0.00001 and 1.0, which of the following is the recommended way to sample a value for α? **1 / 1 point**

- r = np.random.rand()
  alpha = 10**r
- r = -5*np.random.rand()
  alpha = 10**r
- r = -4*np.random.rand()
  alpha = 10**r
- r = np.random.rand()
  alpha = 0.00001 + r*0.99999

Expand

Correct
Yes. This will generate a random value between 10^{-5} and 10^{0} chosen randomly in a logarithmic scale.

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. 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**

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

Expand

Correct

7. When using normalization: **1 / 1 point**

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

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 $z^{(i)}_{\text{norm}} = \frac{z^{(i)}-\mu}{\sqrt{\sigma^2+\epsilon}}$ use of the ϵ parameter prevents that the denominator be 0.

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

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

Expand

Correct

Correct. When applying the linear transformation $\tilde{z}^{(l)} = \beta^{[l]} z^{(l)}_{norm} + \gamma^{[l]}$ 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**

- True
- (selected) False

Expand

Correct

Correct. During the test, the parameters μ and σ^2 are estimated using an exponentially weighted average across mini-batches used during training.

10. Which of the following are some recommended criteria to choose a deep learning framework? **1 / 1 point**

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

Expand

Correct

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