Congratulations! You passed!

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

Retake the assignment in **23h 52m**

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

- False
- True

Expand

Correct

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

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

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.

- [x] mini-batch size

  ✓ **Correct**

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

Correct
Great, you got all the right answers.

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 0.00001 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
- (selected) r = -5*np.random.rand()
  alpha = 10**r
- r = 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**

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

Expand

Correct

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

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

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.
- [ ] $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$.
- [x] The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ set the variance and mean of $\tilde{z}^{[l]}$.

  ✓ **Correct**

  Correct. When applying the linear transformation $\tilde{z}^{(l)} = \beta^{[l]} z^{(l)}_{norm} + \gamma^{[l]}$ we set the variance and mean of $\tilde{z}^{[l]}$.

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

Incorrect
You didn't select all the correct answers

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. 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?

0 / 1 point

- ○ False
- ◉ True

Expand

Incorrect

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