Congratulations! You passed!

Grade received 100% | Latest Submission Grade 100% | 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**

- ☑ α

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

- ☐ β_1, β_2 in Adam.
- ☐ $$\epsilon$$ in Adam.
- ☑ The $$\beta$$ 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**

- ◉ True
- ○ False

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.

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 = -3*np.random.rand()

alpha = 10**r

- r = -5*np.random.rand()
  alpha = 10**r
- r = np.random.rand()
  alpha = 0.001 + r*0.999
- 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? **1 / 1 point**

- (•) True
- ( ) False

Expand

Correct
Correct. 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**

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

Expand

Correct

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

- ( ) There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.
- (•) One intuition behind why batch normalization works is that it helps reduce the internal covariance.
- ( ) The parameter ϵ in the batch normalization formula is used to accelerate the convergence of the model.
- ( ) The parameters β and γ of batch normalization can't be trained using Adam or RMS prop.

Expand

Correct
Yes. Internal covariance is a name to express that there has been a change in the distribution of the activations. Since after each iteration of gradient descent the parameters of a layer change, we might think that the activations suffer from covariance shift.

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 $\beta = \mu$.
- The parameters $z^{[l]}_{norm} = \frac{\gamma^{[l]}}{\sqrt{\sigma^2}}$.
- The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.

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. 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**

- Perform the needed normalizations, use μ and σ^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 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.

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**

- True
- False

Expand

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