coursera

Search in course   Search

Search in course

English ∨   🔔   L ∨

🔔   L ∨

< Previous Next >

≡ Menu

Graded Assignment
Hyperparameter tuning, Batch Normalization, Programming Frameworks

Assignment details

Due	**Attempts**
May 18, 11:59 PM +07May 18, 11:59 PM +07	0 left (3 attempts every 24 hours)
Submitted	**Time limit**
May 13, 5:31 PM +07May 13, 5:31 PM +07	50 minutes per attempt50 min per attempt
	Submissions
	0 left (1 total within the time limit)

↻ Retry

Your grade

To pass you need at least 80%. We keep your highest score.

100%
View submission
See feedback
👍 Like
👎 Dislike
🏳 Report an issue

💬

← Back

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Graded Assignment • 50 min

🌐 English ∨
DueMay 18, 11:59 PM +07

1.

Question 1

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

○ ○

False

◉ ◉

True

1 point

2.

Question 2

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

☑ ☑

ααalpha

☑ ☑

mini-batch size

☑ ☑

The ββbeta parameter of the momentum in gradient descent.

☐ ☐

β_1β1beta, start subscript, 1, end subscript, β_2β2beta, start subscript, 2, end subscript in Adam.

☐ ☐

ϵϵ\epsilon in Adam.
1 point

3.

Question 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?

◉ ◉

False

○ ○

True

1 point

4.

Question 4

Knowing that the hyperparameter ααalpha should be in the range of $0.001$0.0010, point, 001 and $1.0$1.01, point, 0. Which of the following is the recommended way to sample a value for ααalpha?

○ ○

r = np.random.rand()

alpha = 0.001 + r*0.999

○ ○

r = 4*np.random.rand()

alpha = 10**r

◉ ◉

r = -3*np.random.rand()

alpha = 10**r

○ ○

r = -5*np.random.rand()

alpha = 10**r

1 point

5.

Question 5

Finding new values for the hyperparameters, once we have found good ones for a model, should only be done if new hardware or computational power is acquired. True/False?

○ ○

True

◉ ◉

False

1 point

6.

Question 6

When using batch normalization, it is OK to drop the parameter $b^{[l]}$b[l]b, start superscript, open bracket, l, close bracket, end superscript from the forward propagation because it is effectively canceled out during the normalization step, where we compute $z^{[l]}_{norm} = \frac{z^{[l]} - \mu}{\sigma}$znorm[l] = σz[l]−µz, start subscript, start text, n, o, r, m, end text, end subscript, start superscript, open bracket, l, close bracket, end superscript, equals, start fraction, z, start superscript, open bracket, l, close bracket, end superscript, minus, mu, divided by, sigma, end fraction. True/False?

○ ○

False

◉ ◉

True

1 point

7.

Question 7

In the normalization formula $z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$znorm(i) = σ2+ε

z(i)−µz, start subscript, n, o, r, m, end subscript, start superscript, left parenthesis, i, right parenthesis, end superscript, equals, start fraction, z, start superscript, left parenthesis, i, right parenthesis, end superscript, minus, mu, divided by, square root of, sigma, squared, plus, \varepsilon, end square root, end fraction, why do we use epsilon?

○ ○

To speed up convergence

○ ○

To have a more accurate normalization

◉ ◉

To avoid division by zero

○ ○

In case μµmu is too small

1 point

8.

Question 8

Which of the following is true about batch normalization?

○ ○

The optimal values to use for γγgamma and ββbeta are $\gamma = \sqrt{\sigma^2 + \epsilon}$γ = σ2 + ε

gamma, equals, square root of, sigma, squared, plus, \epsilon, end square root and $\beta = \mu$β = µbeta, equals, mu.

○ ○

$z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2}}$znorm(i) = σ2

z(i)−µz, start subscript, n, o, r, m, end subscript, start superscript, left parenthesis, i, right parenthesis, end superscript, equals, start fraction, z, start superscript, left parenthesis, i, right parenthesis, end superscript, minus, mu, divided by, square root of, sigma, squared, end square root, end fraction.

◉ ◉

The parameters $\gamma^{[l]}$γ[l]gamma, start superscript, open bracket, l, close bracket, end superscript and $\beta^{[l]}$β[l]beta, start superscript, open bracket, l, close bracket, end superscript set the variance and mean of $\tilde{z}^{[l]}$z

[l]z, with, \widetilde, on top, start superscript, open bracket, l, close bracket, end superscript.

◯ ○

The parameters $\gamma^{[l]}$γ[l]gamma, start superscript, open bracket, l, close bracket, end superscript and $\beta^{[l]}$β[l]beta, start superscript, open bracket, l, close bracket, end superscript can be learned only using plain gradient descent.

1 point

9.

Question 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?

◯ ○

True

◉ ◉

False

1 point

10.

Question 10

Which of these statements about deep learning programming frameworks are true? (Check all that apply)

☑ ☑

A programming framework allows you to code up deep learning algorithms with typically fewer lines of code than a lower-level language such as Python.

☑ ☑

Even if a project is currently open source, good governance of the project helps ensure that it remains open even in the long term, rather than become closed or modified to benefit only one company.

☐ ☐

Deep learning programming frameworks require cloud-based machines to run.

1 point

Beta