

[Skip to Main Content](#)



SEARCH IN COURSE

Search

• 1



Quang Minh Vũ

1. [Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization](#)
2. [Week 3](#)
3. Hyperparameter tuning, Batch Normalization, Programming Frameworks

PreviousNext

- **Hyperparameter Tuning**
- **Batch Normalization**
- **Multi-class Classification**
- **Introduction to Programming Frameworks**
- **Lecture Notes (Optional)**
- **Quiz**



Quiz: Hyperparameter tuning, Batch Normalization, Programming Frameworks

10 questions

- **Programming Assignment**
- **References & Acknowledgments**

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Quiz 20 minutes • 20 min

Submit your assignment

Due May 2, 1:59 PM +07 May 2, 1:59 PM +07

Attempts 3 every 8 hours

Try again

Receive grade

To Pass 80% or higher

Your grade

80.83%

View Feedback

We keep your highest score

Like

Dislike

Report an issue

[Back](#)

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Graded Quiz • 20 min

Due May 2, 1:59 PM +07

Congratulations! You passed!

Grade received 80.83%

To pass 80% or higher

Go to next item

Hyperparameter tuning, Batch Normalization, Programming Frameworks

Latest Submission Grade 80.83%

1.

Question 1

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

1 / 1 point



False



True

Correct

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

2.

Question 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

☐

β_1, β_2 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.

☐

ϵ in Adam.

☒

α

Correct

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

3.

Question 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 presence of local minima (and saddle points) in your neural network

☐

The number of hyperparameters you have to tune

☒

The amount of computational power you can access

☐

Whether you use batch or mini-batch optimization

Correct

4.

Question 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 β ?

1 / 1 point

☐

$r = \text{np.random.rand}()$ $\beta = r \cdot 0.9 + 0.09$

☒

$r = \text{np.random.rand}()$ $\beta = 1 - 10^{-(r - 1)}$

☐

`r = np.random.rand()` `beta = r*0.09 + 0.9`



`r = np.random.rand()` `beta = 1-10**(- r + 1)`

Correct

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?

0 / 1 point



True



False

Incorrect

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

6.

Question 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 $\tilde{z}^{[l]}_{\text{normalize}} = \beta^{[l]} + \gamma^{[l]} \hat{z}^{[l]} + \gamma^{[l]}$. True/False?

1 / 1 point



True



False

Correct

Correct. 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.

Question 7

In the normalization formula $z_{\text{norm}}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$, why do we use epsilon?

1 / 1 point



In case μ is too small



To have a more accurate normalization



To speed up convergence



To avoid division by zero

Correct

8.

Question 8

Which of the following are true about batch normalization?

0.75 / 1 point



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

Correct

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



$z_{norm}^{[i]} = \frac{z^{[i]} - \mu}{\sqrt{\sigma^2}}$



The parameters $\gamma^{[l]}$ and $\beta^{[l]}$ can be learned only using plain gradient descent.

This should not be selected

No. They can be learned with gradient descent, gradient descent with momentum, RMSprop, and Adam, like all the other parameters.



The optimal values to use for γ and β are $\gamma = \sqrt{\sigma^2 + \epsilon}$ and $\beta = \mu$.

9.

Question 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.



Skip the step where you normalize using μ and σ^2 since a single test example cannot be normalized.



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.



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

Correct

10.

Question 10

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

0.3333333333333333 / 1 point



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

This should not be selected

No. You can run deep learning programming languages from any machine with a CPU or a GPU, either locally or on the cloud.

☐

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.

Correct