

# Session 11 Quiz

---

**Due** Jul 20 at 11:59pm    **Points** 100    **Questions** 12    **Available** after Jul 14 at 4pm    **Time Limit** 45 Minutes

---

## Instructions

Instructions:

1. The context of the questions is what we discussed in the class.
2. You have 45 minutes to attempt the quiz
3. Once you start the quiz, you cannot go back and re-attempt it
4. You will not find answers online, so please make sure you are ready for the quiz
5. For Multiple Answer Questions, ALL the answers must be correct to score any point

All the best!

## Attempt History

	Attempt	Time	Score
LATEST	<a href="#">Attempt 1</a>	23 minutes	55 out of 100

Score for this quiz: **55** out of 100

Submitted Jul 20 at 7:40pm

This attempt took 23 minutes.

**Question 1**

**7.5 / 10 pts**

Select which all are true:

Correct!

☒ As the loss value reduces, we should reduce the learning rate as well

Correct!

☒ Realistically, we can never find global minima for any DNN

Correct!

☒ Two networks with different batch sizes will have different loss values after 1st epoch

Correct Answer

☐ If the ground truth labels are wrong, you will always see overfitting

## Question 2

7.5 / 10 pts

Which of the following are true?

Correct Answer

☐ While updating the weights of one kernel, we must assume other kernel is constant

Correct!

☒ For gradient descent we take negative value of the gradients

Correct!

☒ A high dropout value can lead to lot of fluctuations at the later stages of training in VA

Correct!

☒ Comparatively, at the beginning of the training, learning rate can be much higher than at later stages

**Question 3****5 / 10 pts**

Which of the following are true?

**Correct!**

☒ As the VA increases, we should increase the batch size keeping the learning rate constant

**Correct Answer**

☐ As the VA increases, we should reduce the learning rate, keeping batch size constant

**Correct!**

☒ In SDG (academically), batch size is 1

**Correct Answer**

☐ Even if we use momentum with SGD, the learning rate remains constant.

**Question 4****0 / 10 pts**

It is proven that if we add gradient perturbation (small noise in gradients), we can avoid hitting the problem of weights getting stuck in plateaus.

Which of the following can have a similar effect:

☐ L1/L2 regularization

**Correct Answer**

☐ Patch Gaussian

**You Answered**

☒ ReLU

**Correct!**☒ Dropout**Question 5****10 / 10 pts**

Match the following: Left is problem, right is solution

**Correct!****Weight Plateaus**

Gaussian Noise ▼

**Correct!****UnderFitting**

Remove Dropout ▼

**Correct!****OverFitting**

Image Augmentation ▼

**Correct!****Slow Convergence**

Increase LR ▼

Other Incorrect Match Options:

- SGD
- Reduce LR
- CutOut

**Question 6****5 / 5 pts**

Assume that the value of a specific weight was 4.

The derivative of the Loss Function w.r.t. this weight is 100

If we used a learning rate of 0.01, after the backprop step, what would be the value of the new weight?

**Correct!****Correct Answers**

3.0

3

**Question 7****0 / 5 pts**

We are working on a custom dataset, where we have 10 classes, but only 100 images for each class. Which Optimization Algorithm should provide us better results?

**Correct Answer**☐ Adam**You Answered**☒ SGD with Momentum

- ☐ SGD
- ☐ SGD with reducing learning rates

**Question 8****0 / 5 pts**

In the momentum algorithm, what would be the value of the  $v^{(t)}$  for the very first time?



- ☐ would be set to a random value
- ☒ Would be required to be calculated

**You Answered****Correct Answer**

- ☐ 0

**Question 9****5 / 15 pts**

What all would be the benefits of adding momentum term to SGD?

**Correct Answer**

- ☐ Faster Convergence

**Correct!**☒ Solving weight plateauing problem**Correct Answer**☐ Solving weight saddling problem**Question 10****5 / 10 pts**

We "know" for sure that we are stuck in local minima. What all could we try?

**Correct!**☒ Add momentum to our optimizer if we haven't done so yet**Correct Answer**☐ Change the optimizer for sometime**Correct!**☒ Add image augmentation**Correct Answer**☐ Increase the learning rate**Question 11****5 / 5 pts**

We defined a network and then ran an LR finder on it.

After a few tests on learning rates, we do not see any change in the loss function. What all could be wrong?

**Correct!**☒ Network is incapable of learning**Correct!**☒ We haven't yet tried varied range of possible LRs**Question 12****5 / 5 pts**

Match the following

**Correct!****SGD with Momentum**

Most DNNs ▼

**Correct!****Adaptive Optimizers**

GANs &amp; RL ▼

Quiz Score: **55** out of 100