

Q10

Due Apr 1 at 5:30am**Points** 200**Questions** 11**Available** Mar 18 at 9:30am - Apr 1 at 5:30am 14 days**Time Limit** 30 Minutes

Instructions

You have 30 minutes to solve the quiz.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	19 minutes	170 out of 200

Score for this quiz: **170** out of 200

Submitted Mar 31 at 7:47pm

This attempt took 19 minutes.

Question 1

20 / 20 pts

Which of the following are true?

Correct!

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

Correct!

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

Correct!

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

Correct!

- ☒ For gradient ascent we take positive value of the gradients

Question 2**15 / 20 pts**

Which of the following are true?

Correct!

- ☒ In SDG (academically), batch size is 1

Correct!

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

Incorrect Answer

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

Correct!

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

Question 3**10 / 20 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:

Correct!

- ☒ Patch Gaussian

☐ L1/L2 regularization

Incorrect Answer

☐ Dropout

☐ ReLU**Question 4****15 / 20 pts**

Match the following: Left is the problem, right is the solution

Correct!**Weight Plateaus**

Gaussian Noise ▼

Correct!**UnderFitting**

Remove Dropout ▼

Correct!**OverFitting**

Image Augmentation ▼

Not Answered**Slow Convergence**

SGD ▼

Correct Answer

Increase LR

Other Incorrect Match Options:

- CutOut
- Reduce LR
- SGD

Question 5**20 / 20 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!

3

Correct Answers

3

3.0

Question 6

10 / 10 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?

☐ SGD

☐ SGD with reducing learning rates

☐ SGD with Momentum

Correct!

☒ Adam

Question 7

10 / 10 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

Correct!☒ 0**Question 8****20 / 20 pts**

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

Correct!☒ Faster Convergence**Correct!**☒ Solving weight plateauing problem**Correct!**☒ Solving weight saddling problem**Question 9****10 / 20 pts**

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

Correct!☒ Increase the learning rate**Incorrect Answer**☐ Add image augmentation**Incorrect Answer**☐ Change the optimizer for sometime**Correct!**☒ Add momentum to our optimizer if we haven't done so yet**Question 10****20 / 20 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!

☒ We haven't yet tried varied range of possible LR's

Correct!

☒ Network is incapable of learning

Question 11

20 / 20 pts

Match the following

Correct!

SGD with Momentum

Most DNNs ▼

Correct!

Adaptive Optimizers

GANs & RL ▼

Other Incorrect Match Options:

- NLP
- Object Detection Networks

Quiz Score: **170** out of 200