# Session 4 Quiz

**Due** Dec 9, 2019 at 5:59am **Points** 100 **Questions** 10 **Available** until Dec 9, 2019 at 5:59am
**Time Limit** 36 Minutes

## Instructions

Instructions:

1. You have 36 minutes to attempt the quiz
2. Once you start the quiz, you cannot go back and re-attempt it
3. You will not find answers online, so please make sure you are ready for the quiz
4. For Multiple Answer Questions, ALL the answers must be correct to score any point

This quiz was locked Dec 9, 2019 at 5:59am.

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 6 minutes | 58.33 out of 100 |

Score for this quiz: **58.33** out of 100
Submitted Dec 9, 2019 at 12:21am
This attempt took 6 minutes.

### Question 1

**15 / 20 pts**

According to the universal approximation theorem, given enough capacity, we know that a feedforward network with a single layer is sufficient to represent any function. However, the layer might be massive and the network is prone to overfitting the data. Therefore, there is a common trend in the research community that our network architecture needs to go deeper.

Deeper layers help in:

Correct! ☑ Achieving higher receptive fields

Correct! ☑ Break down large objects in smaller components

Correct Answer ☐ Easier training when compared to only 1 layer

Correct! ☑ Creating large number of correlation templates for same objects

### Question 2

**5 / 10 pts**

The receptive field of a network is 128, but the images it is trained on is only 32. What is going on?

Correct Answer ☐ Large receptive field is making a large template with many orientations/object-variants for better correlation/convolution

Correct!

- [x] Large receptive field is making templates which include background for the objects as well
- [ ] Large Receptive fields add many parameters because of which network can learn better
- [ ] Large Receptive fields are made through deeper networks which is always a good idea.

## Question 3

0 / 5 pts

If we add 16 Identity Layers to VGG16, what would happen to the performance? (Identity layers don't do anything)

Correct Answer

- [ ] Will remain same

You Answered

- [x] Will increase
- [ ] Will reduce
- [ ] Will reduce, but can be trained to increase

## Question 4

0 / 5 pts

Does ResNet suffers from Checkerboard issue?

Correct Answer

- ( ) True

You Answered

- (•) False

## Question 5

5 / 5 pts

If every Identity mappings in the ResNet50 architecture is multiplied by 0.5, what will happen?

Correct!

- (•) Vanishing Gradient
- ( ) Nothing
- ( ) Exploding Gradients
- ( ) Lesser accurate network, since only half the actual amplitudes of channels can pass through

## Question 6

0 / 5 pts

If all the 3x3 stride 2 kernels in ResNet architecture are replaced by 3x3 stride followed by Maxpooling, do you think the accuracy will increase or reduce?

Correct Answer
- Increase
- Reduce
- Can't Say

You Answered
- It is a secret and I don't want to share it with anyone

## Question 7

13.33 / 20 pts

A robust minima will

Correct Answer
- be more like a plateau than being spikey
- be more like a spikey than being plateau

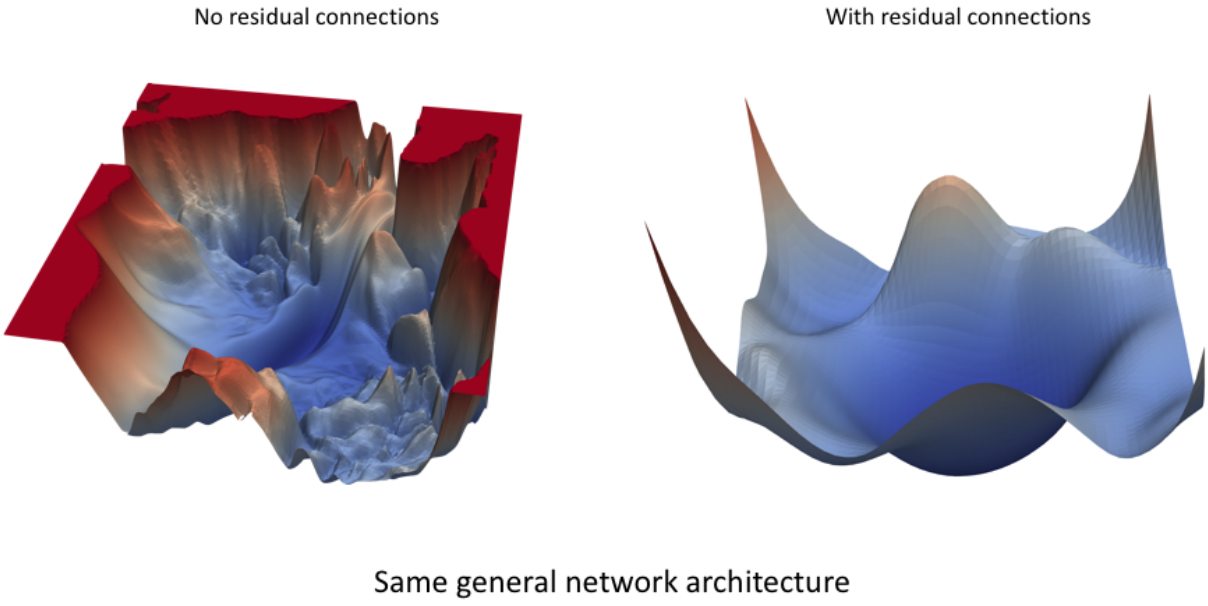
Correct!
- not necessarily would be the global minima

Correct!
- will make the model generalize better to unseen data

## Question 8

0 / 5 pts

Which loss topology from the image below would you prefer, left or right?



CANNOT SEE THE IMAGE? ONLY IS YOU CANNOT SEE THE IMAGE ABOVE, OPEN THIS (https://www.jeremyjordan.me/content/images/2018/02/Screen-Shot-2018-02-26-at-10.50.53-PM.png) IMAGE IN ANOTHER TAB (RIGHT CLICK)

Correct!
- [x] Right one, as that has smoother loss topology

- [ ] Right one, as that looks great

You Answered
- [x] Right one, as it looks like it has lower loss value

- [ ] Right one as that has more number of minimas

## Question 9

0 / 5 pts

Smaller datasets and architectures seem to require larger values for weight decay while larger datasets and deeper architectures seem to require smaller values.

Correct Answer
- [ ] One hypothesis could be that complex data providesits own regularization and other regularization should be reduced

You Answered
- [x] One hypothesis could be that complex data need parameters to have smaller values, hence smaller weight decay is required

## Question 10

20 / 20 pts

Regularization is the process of introducing additional information in order to solve ill-posed problems or prevent over-fitting. A trivial example is when trying to fit a simple linear regression but you only have one point. In this case, you can't estimate both the slope and intercept (you need at least two points) so any estimate (which *only* uses the data) will be ill-formed. Instead, if you provide some "additional information", you can get a much more reasonable estimate.

Using this exact definition and its extension, what from below can be considered regularization

Correct!
- [x] DropOut

Correct!
- [x] L1 & L2 weight decay

Correct!
- [x] Image Augmentation

Correct!
- [x] Small Gaussian Noise to the Loss function

Correct!
- [x] Large Learning Rate stages in OneCycleLR strategy

Quiz Score: **58.33** out of 100