ResNet

Paper 1: Deep Residual Learning for Image Recognition

Paper 2: Identity Mapping in Deep Residual Networks

1. Can stacking more layers will give you a better network? What is the degradation problem discussed in paper 1?

2. How the author concludes that the degradation problem is not caused by over-fitting? Explain

3. What should be the behavior of a model with optimal identity functions in ResNet setup?

4. Explain all the methods that can be used to add a residual connection with (a) same dimensions and (b) different dimensions.

5. Explain how the ResNet model discussed in paper 1 achieves a deeper model with a smaller number of FLOPs as compared with VGG Net.

6. Explain the properties exhibited by equation (4) in paper 2.

7. Name all the types of shortcut connections used in paper 2. Which method of connection is best (as per paper 2) and explain why?

8. Why does optimization face difficulty in ResNets when the shortcut signal is scaled down?

9. Can ResNet architecture helps achieve better accuracy when the model is not overly deep? Explain why or why not?

10. Describe briefly. (a) Residual building block, (b) Solution space (for classification problem), (c) bottleneck architecture, (d) pre-activation and post-activation