Deep Learning

6.1 Going Deeper: Benefits of depth

Dr. Konda Reddy Mopuri
kmopuri@iittp.ac.in
Dept. of CSE, IIT Tirupati

Depth and performance

1. Deeper architectures improved performance in many applications (e.g. object recognition)

Depth and performance

1. Deeper architectures improved performance in many applications (e.g. object recognition)
2. LeNet (8), AlexNet (8), VGG (16, 19), GoogLeNet (22, . . . , 76), ResNet (34, . . . , 152)

Depth and Performance

model	top-1 err.	top-5 err.
VGG-16 [41]	28.07	9.33
GoogLeNet [44]	-	9.15
PReLU-net [13]	24.27	7.38
plain-34	28.54	10.02
ResNet-34 A	25.03	7.76
ResNet-34 B	24.52	7.46
ResNet-34 C	24.19	7.40
ResNet-50	22.85	6.71
ResNet-101	21.75	6.05
ResNet-152	**21.43**	**5.71**

Figure credits: He et al. 2015

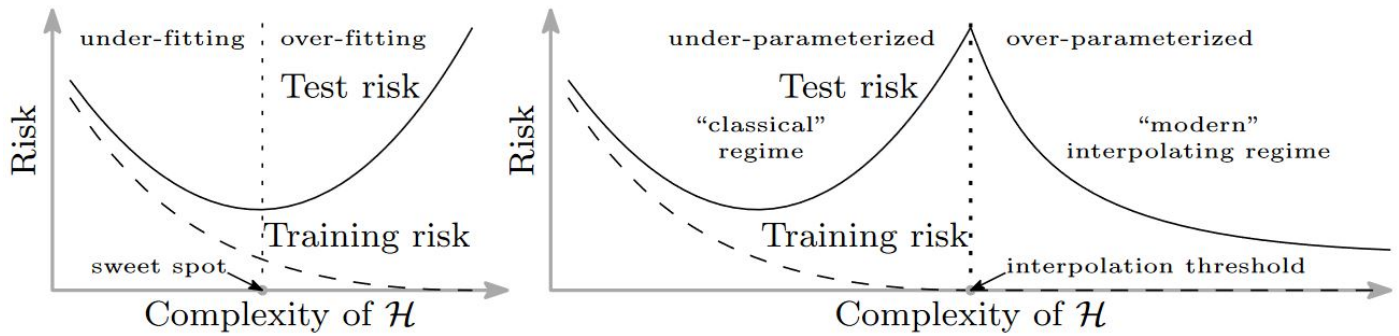
Complexity vs. depth and width

1. A measure of complexity of the mapping learned by the DNN increases
   - exponentially with its depth
   - linearly with the layers' width

Telgarsky 2015, 2016

Depth and Over-fitting

1. Contrary to what we have discussed in Bias-Variance decomposition, Over parameterizing a DNN often improves the performance



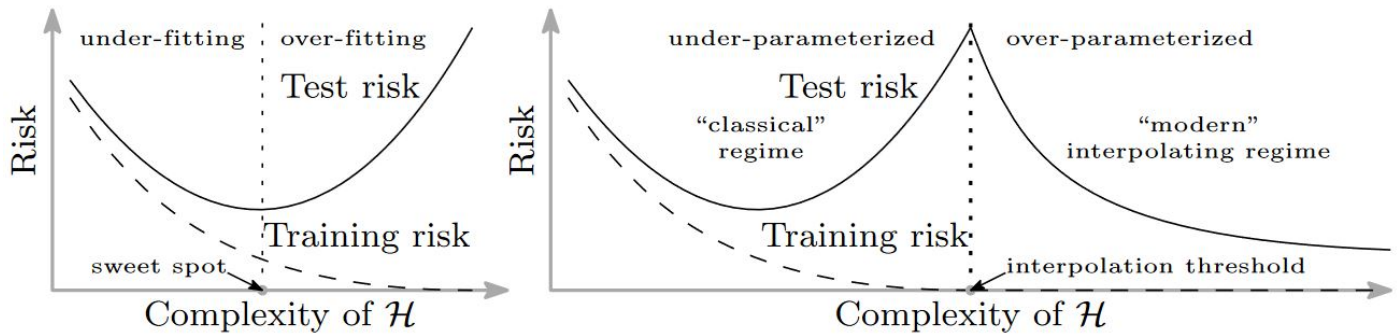
(a) U-shaped "bias-variance" risk curve

(b) "double descent" risk curve

Belkin et al. 2018

Depth and Over-fitting

1. Contrary to what we have discussed in Bias-Variance decomposition, Over parameterizing a DNN often improves the performance
2. Only bias component (along with regularization) drives the optimization



(a) U-shaped "bias-variance" risk curve

(b) "double descent" risk curve

Belkin et al. 2018

Depth and Gradient

1. We now have a good reason to develop deeper models

Depth and Gradient

1. We now have a good reason to develop deeper models
2. However, amplitude of gradients also needs to be controlled

Depth and Gradient

1. We now have a good reason to develop deeper models
2. However, amplitude of gradients also needs to be controlled
3. Gradients should not vanish

Depth and Gradient

1. We now have a good reason to develop deeper models
2. However, amplitude of gradients also needs to be controlled
3. Gradients should not vanish
4. Gradients should be homogeneous at all the layers

Gradient and Depth

1. Because of our concern to take care of gradients, we often compromise on the family of functions learned by the DNN architectures