GENERALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Since September 2019, I was reading about Robustness+ Network Compression. The more I am reading, the more I get that there is a strong correlation between robustness/stability, generalization, generalization bounds, and compression. I found some interesting papers and talks regarding this and I will cover them here.

GENERALIZATION AND COMPRESSION

The whole story starts with figure **??** . While we expect that with more parameters in model we have less generalization, in Deep Neural Network it is not the case.

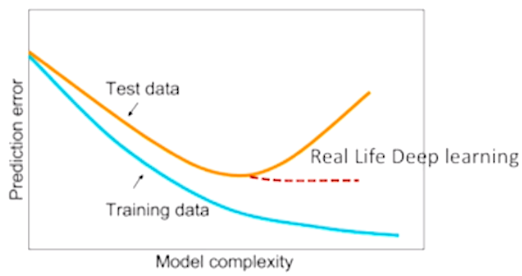

Figure 1: Generalizaion gets better with overparametrized deep neural networks(Arora et al. (2018))

In order to solve this mystery we might have a look on generalization bounds in ML:

$$error_{test} \leq error_{training} + c\sqrt{\frac{capacity}{m}} \tag{1}$$

In which, the capacity roughly corresponds to the number of parameters and m is the number of training samples. However equation 1 is not working for DL because the number of params are way more than m(very loose bound). Many other bounds proposed but their common problem is that they are way more than number of params.

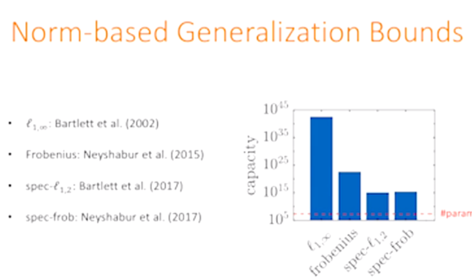

Figure 2: Norm based generalization bounds are way more than number of parameters (Arora et al. (2018))

Arora et al. (2018) tries to answer this question: What property of network trained on real data can help to sharpen this bound? They defined noise stability of trained neural network: How injected gaussian noise at a layer affects higher layers? *i.e.*, noise propagation in layers.

References

Sanjeev Arora, Rong Ge, Behnam Neyshabur, and Yi Zhang. Stronger generalization bounds for deep nets via a compression approach. *arXiv preprint arXiv:1802.05296*, 2018.

Appendix

You may include other additional sections here.