Generalization Bounds for Rank-sparse Neural Networks: Additional Experiments on CIFAR with CNNs

Anonymous Authors[1]

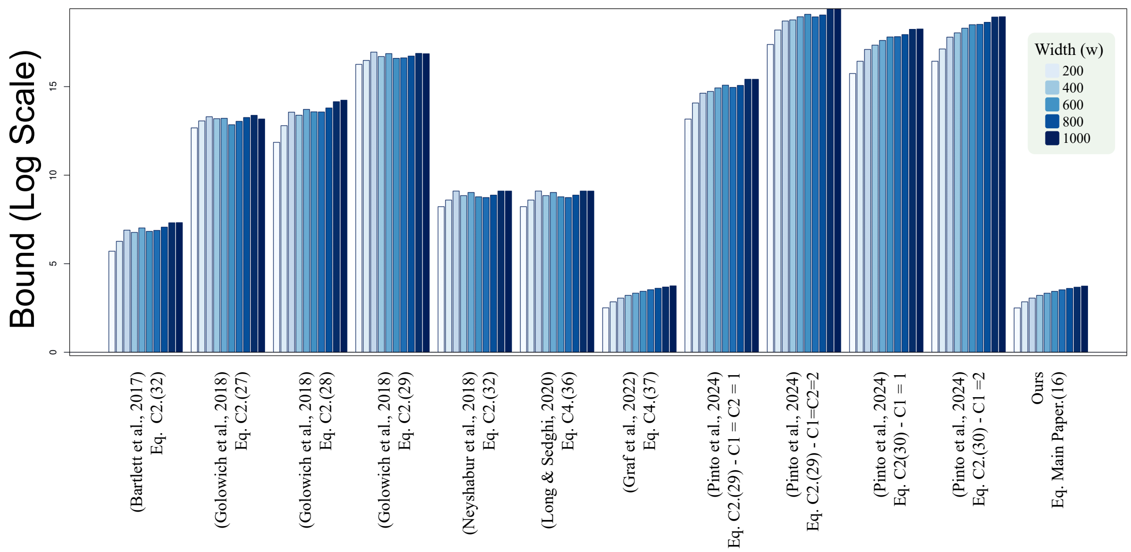

Figure 1: Numerical Comparison of Bounds with Related Works for the CIFAR10 Dataset

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

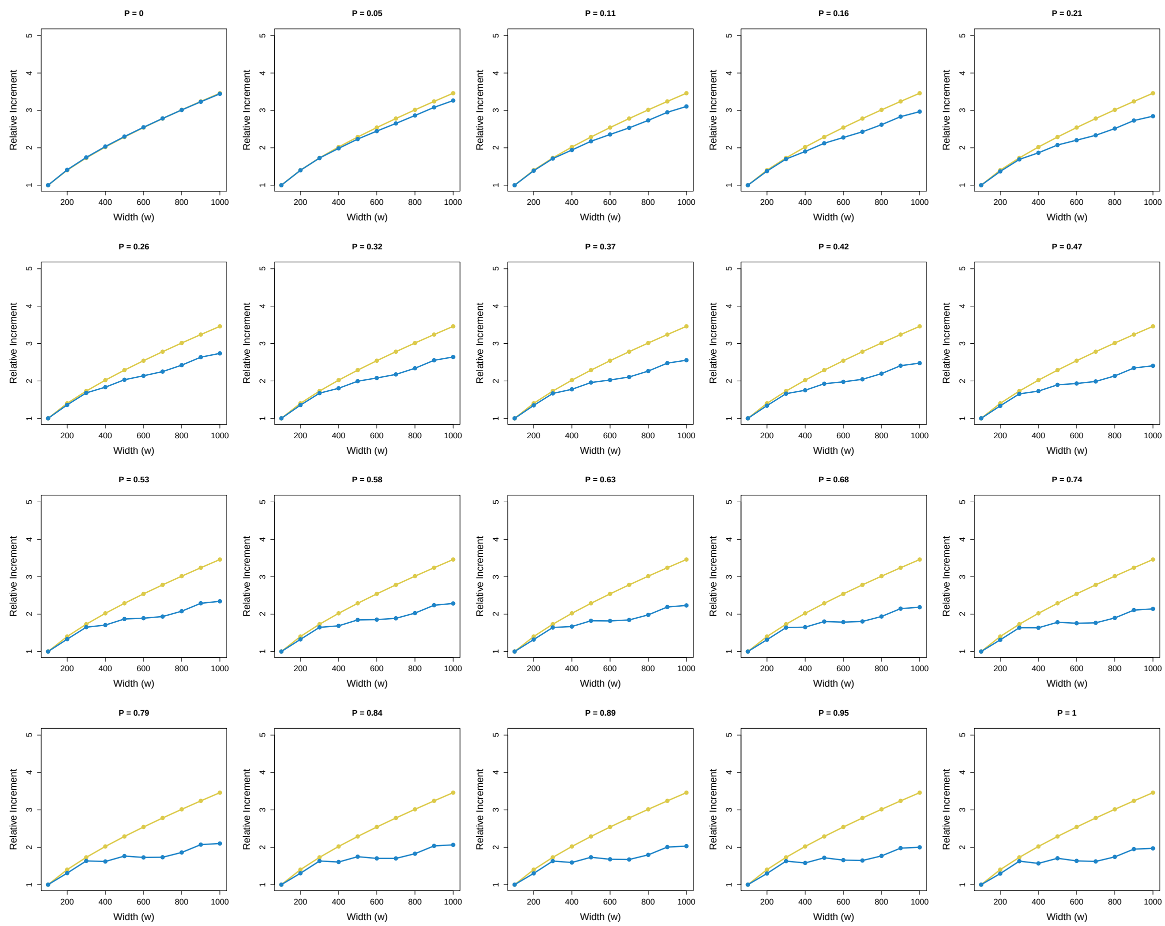

Figure 2: Comparison of Relative Trends in Graf et al. (Equation C4. (37)) and our Work for Various p_ℓ. The graphs are normalized so the bound with width 100 is 1. Although numerically, smaller values of p yield tighter bounds due to a milder dependence on scaling factors and Lipschitz constants, moderate values of p exhibit milder growth relative to w, which better aligns with the true generalization behavior of the model.

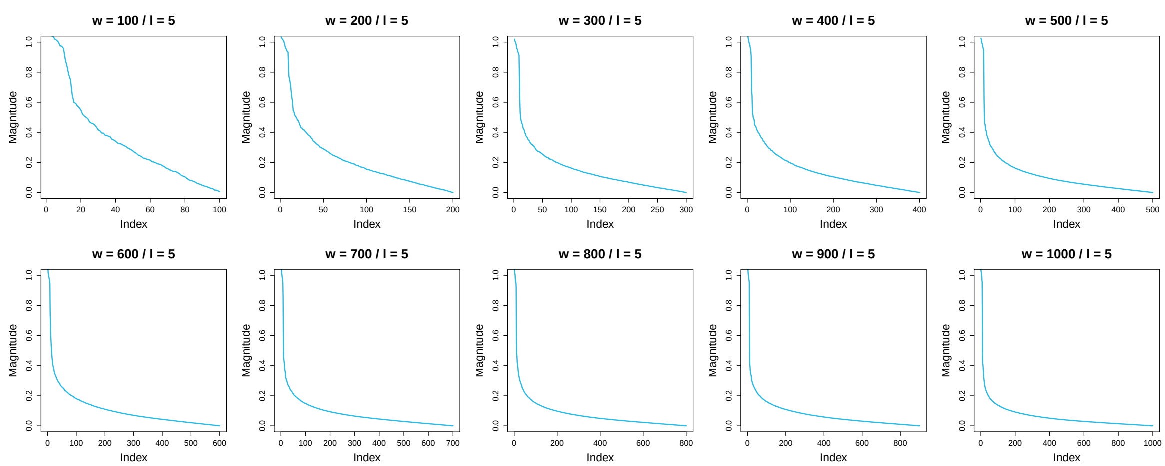

Figure 3: Singular Spectrum of Second Fully Connected Layer for Different Widths: we see a sharper spectral decay for larger values of the width, consistent with the hypothesis that a minimum bottleneck rank is required to represent the data.

3