IMPROVED GENERALIZATION BOUNDS FOR DEEP NEURAL NETWORKS USING GEOMETRIC FUNCTIONAL ANALYSIS

Anonymous authors
Paper under double-blind review

A APPENDIX

A.1 THEOREM PROOFS

Theorem 4. *The product of k and ρ is bounded by the metric (n, d)-span of the set Z for any n and d_l*

$$\omega_{d_l}(Z) \geq k\rho \tag{1}$$

Proof. We will recall the definitions of $C_i(n, d)$ and $M_d(\epsilon)$ to derive a bound on $M_d(\epsilon)$

$$\begin{aligned} C_i(n,d) &\triangleq 2^i \binom{n}{i} (d-i)^i \\ M_d(\epsilon) &\triangleq \sum_{i=0}^{n-1} C_i(n,d) \left(\frac{1}{\epsilon}\right)^i \end{aligned} \tag{2}$$

from these equations it is easy to see that

$$M_d(\epsilon) \leq \frac{(4d)^n}{\epsilon^{n-1}} \tag{3}$$

because ρ is the minimal distance between any 2 points of the finite set Z, and cardinality of Z is k, as the minimal distance between any two points is ρ, the covering number $M(\epsilon, Z)$ is k, because around every point we draw a ball of radius ϵ and since these balls need not be disjoint, the union of these balls covers the entire set Z.So for the set k, we have $M(\epsilon, Z) = k$.

substituting equation (3) in the below equation

$$\omega_d(Z) = \sup_{\epsilon \geq 0} \epsilon^n [M(\epsilon, Z) - M_d(\epsilon)] \tag{4}$$

we get the final result as

$$\begin{aligned} \omega_d(Z) &\geq [k\rho^n - (4d)^n \rho] \\ &\geq k\rho \end{aligned} \tag{5}$$

□

A.2 EXPERIMENTAL SETUP

In all the experiments we use SGD with learning rate 0.1 and batch size 32. We train the network on a training dataset of 60000 images of flatten to a input vector of size 784. We stop training when we classify at least 99% of the data perfectly and ρ is computed by randomly sampling some 1000 datapoints from the training dataset.