

# **TTIC 31230, Fundamentals of Deep Learning**

David McAllester, Autumn 2020

## **Learning Theory II**

### **The PAC-Bayes Guarantee**

## The PAC-Bayes Guarantee

Let  $p$  be any “prior” and  $q$  be any “posterior” on any (possibly continuous) model space. Define

$$L(q) = E_{h \sim q} L(h)$$

$$\hat{L}(q) = E_{h \sim q} \hat{L}(h)$$

For any  $p$  and any  $\lambda > \frac{1}{2}$ , with probability at least  $1 - \delta$  over the draw of the training data, the following holds simultaneously for all  $q$ .

$$L(q) \leq \frac{1}{1 - \frac{1}{2\lambda}} \left( \hat{L}(q) + \frac{\lambda L_{\max}}{N_{\text{Train}}} \left( K L(q, p) + \ln \frac{1}{\delta} \right) \right)$$

## Adding Noise Simulates Limiting Precision

Assume  $0 \leq \mathcal{L}(\Phi, x, y) \leq L_{\max}$ .

Define:

$$\mathcal{L}_\sigma(\Phi) = E_{(x,y) \sim \text{Pop}, \epsilon \sim \mathcal{N}(0,\sigma)^d} \mathcal{L}(\Phi + \epsilon, x, y)$$

$$\hat{\mathcal{L}}_\sigma(\Phi) = E_{(x,y) \sim \text{Train}, \epsilon \sim \mathcal{N}(0,\sigma)^d} \mathcal{L}(\Phi + \epsilon, x, y)$$

Theorem: With probability at least  $1 - \delta$  over the draw of training data the following holds **simultaneously** for all  $\Phi$ .

$$\mathcal{L}_\sigma(\Phi) \leq \frac{10}{9} \left( \hat{\mathcal{L}}_\sigma(\Phi) + \frac{5L_{\max}}{N_{\text{Train}}} \left( \frac{\|\Phi - \Phi_{\text{init}}\|^2}{2\sigma^2} + \ln \frac{1}{\delta} \right) \right)$$

# Non-Vacuous Generalization Guarantees

Model compression has recently been used to achieve “non-vacuous” PAC-Bayes generalization guarantees for ImageNet classification — error rate guarantees less than 1.

Non-Vacuous PAC-Bayes Bounds at ImageNet Scale.

Wenda Zhou, Victor Veitch, Morgane Austern, Ryan P. Adams,  
Peter Orbanz

ICLR 2019

**END**