TTIC 31230 Fundamentals of Deep Learning

Regularization and Generalization Problems

Problem 1. The Stationary Points for L_2 Regularization. Consider the regularized objective

$$\Phi^* = \operatorname*{argmin}_{\Phi} \ E_{(x,y)\sim\text{Train}} \ \left(\mathcal{L}(\Phi, x, y) + \frac{1}{2N_{\text{train}}\sigma^2}||\Phi||^2 \right)$$

By setting the gradient of the objective to zero, solve for Φ as a function of the average gradient g defined by

$$g = E_{\langle x,y \rangle \sim \text{Train}} \nabla \Phi \mathcal{L}(\Phi, x, y).$$

Solution:

$$\nabla_\Phi E_{(x,y)\sim\text{Train}} \ \mathcal{L}(\Phi, x, y) + \frac{1}{2N_{\text{train}}\sigma^2}||\Phi||^2$$

$$= \ \left(E_{(x,y)\sim\text{Train}} \ \nabla_\Phi \mathcal{L}(\Phi, x, y) \right) + \frac{1}{N_{\text{train}}\sigma^2}\Phi$$

$$= \ g + \frac{1}{N_{\text{train}}\sigma^2}\Phi \ = 0$$

$$\Phi \ = \ -N_{\text{train}}\sigma^2 g$$

Note that a larger sample size justifies having a larger norm for the parameter vector.