

Lecture 24: Deep neural networks

Professor Ilias Bilonis

The minimization of the loss function as a stochastic optimization problem

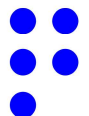
What is a stochastic optimization problem?

$$\min_{\theta} \mathbb{E}_{\mathbf{z}} [\underbrace{l(\theta; \mathbf{z})}_{\text{objective function}}]$$

parameters random vector

||
V

Stochastic optimization



The loss minimization as a stochastic optimization problem

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i; \theta)]^2$$

$$\min_{\theta} L(\theta)$$

$$\min_{\theta} \mathbb{E}[L(\theta; \mathbf{z})]$$

Let $\mathbf{I} \sim \text{Categorical}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$; $\mathbf{I} = \begin{cases} 1, & \text{w/ prob. } \frac{1}{n} \\ 2, & \text{w/ prob. } \frac{1}{n} \\ \vdots & \text{w/ prob. } \frac{1}{n} \end{cases}$

$$l(\theta; \mathbf{I}) = [y_{\mathbf{I}} - f(x_{\mathbf{I}}; \theta)]^2$$

$$\mathbb{E}_{\mathbf{I}}[l(\theta; \mathbf{I})] = \sum_{i=1}^n p(\mathbf{I}=i) [y_i - f(x_i; \theta)]^2$$

$$= \frac{1}{n} \sum_{i=1}^n [y_i - f(x_i; \theta)]^2 = L(\theta)$$



→ Let m to be an integer in $\{1, \dots, n\}$ (batch size)

→ $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_m$ iid. $\text{Categorical}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$

$$\ell_m(\theta; \underbrace{\mathbf{I}_{1:m}}_{\mathbf{z}}) = \frac{1}{m} \sum_{j=1}^m [y_{\mathbf{I}_j} - f(x_{\mathbf{I}_j}; \theta)]^2$$

$$\mathbb{E}_{\mathbf{I}_{1:m}}[\ell_m(\theta; \mathbf{I}_{1:m})] = \mathbb{E}\left[\frac{1}{m} \sum_{j=1}^m [y_{\mathbf{I}_j} - f(x_{\mathbf{I}_j}; \theta)]^2\right]$$

$$= \frac{1}{m} \sum_{j=1}^m \mathbb{E}[y_{\mathbf{I}_j} - f(x_{\mathbf{I}_j}; \theta)]^2$$

$$= \frac{1}{m} \cdot m L(\theta) = L(\theta)$$