

Lecture 24: Deep neural networks

Professor Ilias Bilonis

The minimization of the loss function as a stochastic optimization problem

Loss minimization \rightarrow Stochastic optimization

What is a stochastic optimization problem?

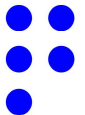
$$\min_{\theta} \mathbb{E}_{\mathbf{z}} [\underbrace{l(\theta; \mathbf{z})}_{\text{objective function}}]$$

parameters (pointing to θ)
random vector (pointing to \mathbf{z})



Stochastic optimization
algorithms

(tend to behave better
than deterministic optimization)



The loss minimization as a stochastic optimization problem

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

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

equivalent $\min_{\theta} \mathbb{E}[l(\theta; \mathcal{I})]$ - stochastic

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

r.v.

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

random index

$$\mathbb{E}[l(\theta; \mathcal{I})] \stackrel{\text{def.}}{=} \sum_{i=1}^n p(\mathcal{I}=i) [y_i - f(x_i; \theta)]^2$$

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



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

$\rightarrow I_1, I_2, \dots, I_m$ iid. $\text{Categorical}(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$

$\rightarrow l_m(\theta; \underline{\mathcal{I}_{1:m}}) = \frac{1}{m} \sum_{j=1}^m [y_{\mathcal{I}_j} - f(x_{\mathcal{I}_j}; \theta)]^2$

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

$L(\theta)$

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

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

$1 \leq m \leq n$
observations