Lecture 24: Deep neural networks

Professor Ilias Bilionis

The minimization of the loss function as a stochastic optimization problem

Loss minimization → Stochastic optimization

PREDICTIVE
SCIENCE LABORATORY

What is a stochastic optimization problem?

parameters

$$\min_{\theta} \mathbb{E}_z\left[\ell(\theta; z)\right]$$

objective function

random vector

$\Downarrow$

Stochastic optimization algorithms

(tend to behave better than deterministic optimization

The loss minimization as a stochastic optimization problem

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

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

$$\Updownarrow \text{ equivalent}$$

$$\min_{\theta} \mathbb{E}[\ell(\theta;Z)] \text{ - stochastic}$$

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

$$\ell(\theta;I) = \left[y_I - f(x_{\underbrace{I}_{\text{random index}}};\theta)\right]^2$$

$$\mathbb{E}_I[\ell(\theta;I)] \overset{\text{def.}}{=} \sum_{i=1}^{n} \underbrace{p(I=i)}_{1/n}\left[y_i - f(x_i;\theta)\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(x_i;\theta)\right]^2 = L(\theta)$$

PREDICTIVE SCIENCE LABORATORY

10

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

→ $I_1, I_2, \dots, I_m$ i.i.d. $\text{Categorical}(1/n, 1/n, \dots, 1/n)$

→ $$\ell_m(\theta; \underbrace{I_{1:m}}_{Z}) = \frac{1}{m}\sum_{j=1}^{m}\left[y_{I_j} - f(x_{I_j};\theta)\right]^2$$

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

$$= \frac{1}{m}\sum_{j=1}^{m}\underbrace{\mathbb{E}\left[\{y_{I_j} - f(x_{I_j};\theta)\}^2\right]}_{L(\theta)}$$

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