Deep Learning

Theoretical Exercises – Week 10 – Chapter 8

Exercises on the book ”Deep Learning” written by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
Exercises and solutions by T. Méndez and G. Schuster

FS 2024

1 Exercises on Optimization for Training Deep Models

1. What is the differences between learning and pure optimization?
2. Why is gradient descent better suited for training neural networks than Newton's method?
3. Assume you have a function

$$g(\boldsymbol{x}) = g(x_1, x_2) = \begin{cases} 1, & x_1^2 + x_2^2 < 1 \text{ and } x_1, x_2 \geq 0 \\ 0, & \text{otherwise} \end{cases},$$

which is defined on the two-dimensional random variable $\boldsymbol{x} = [x_1,\ x_2]^T$ with the following probability density function

$$p(\boldsymbol{x}) = p(x_1, x_2) = \begin{cases} 1, & 0 \leq x_1, x_2 < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Now, from this function $g(\boldsymbol{x})$ the expected value $\mathbb{E}_{\mathrm{x}}[g(\boldsymbol{x})]$ has to be calculated using two different methods:

(a) With the definition of the expected value, thus, by calculating the following integral

$$\mathbb{E}_{\mathrm{x}}[g(\boldsymbol{x})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \cdot g(x_1, x_2)\, dx_1\, dx_2$$

(b) With the Monte Carlo approach, by randomly taking samples $\boldsymbol{x_i}$ according to the probability density function $p(\boldsymbol{x})$ and determining their function value $g(\boldsymbol{x}_i)$. The expected value can then be estimated as

$$\hat{\mathbb{E}}_{\mathrm{x}}[g(\boldsymbol{x})] = \frac{1}{M} \sum_{i=1}^{M} g(\boldsymbol{x}_i),$$

where M is the number of samples. The estimated expected value $\hat{\mathbb{E}}_{\mathrm{x}}[g(\boldsymbol{x})]$ converges to the true expected value $\mathbb{E}_{\mathrm{x}}[g(\boldsymbol{x})]$ from exercise (a), as the number of samples M goes to infinity.