Deep Learning

Theoretical Exercises – Week 5 – Chapter 5

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 Machine Learning Basics

1. Given is a set of samples $\{x^{(1)}..., x^{(m)}\}$ that are independently and identically distributed according to a uniform distribution on the interval $[-0.8, 1.2]$. In the last series of exercises it was shown that the sample mean

$$\hat{\mu}_m = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

is an unbiased estimator and has an expected value of

$$\mathbb{E}[\hat{\mu}_m] = \mu = 0.2.$$

Now check if the estimator also is consistent (variance goes to zero as the number of samples goes to infinity.).

2. Given is one samples $\{x^{(1)}\}$ of a Gaussian distribution with a variance of $\sigma^2 = 1$ and a unknown mean value.

   (a) Determine the maximum likelihood estimation of the mean.

   (b) Repeat the exercise with a set of two samples $\{x^{(1)}, x^{(2)}\}$.

3. Why is the cross-entropy minimized when training a neural network?

4. What is the difference between gradient descent and stochastic gradient descent and why is stochastic gradient descent used for deep learning?

5. Why is it "allowed" to use stochastic gradient descent instead of gradient descent (why does it still work)?