Deep Learning

Theoretical Exercises – Week 4 – 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

Hint:
Several answers are correct in the multiple choice exercises.

1. The goal of machine learning is to achieve ...
   - [ ] ... a small training error.
   - [ ] ... a large training error.
   - [x] ... a small test error.
   - [ ] ... a large test error.
   - [x] ... a small generalization error.
     The test error is an estimation of the generalization error.
   - [ ] ... a large generalization error.
2. An overfitted model has ...
   - [x] ... a large test error.
   - [ ] ... a small test error.
   - [ ] ... a large training error.
   - [x] ... a small training error.
3. An underfitted model has ...
   - [x] ... a large test error.
   - [ ] ... a small test error.
   - [x] ... a large training error.
   - [ ] ... a small training error.

4. A model tends to overfit when ...
   - [x] ... the training set is small.
   - [x] ... the regularization term has little weight.
   - [ ] ... the capacity is smaller than the complexity of the task.
   - [ ] ... the test error is close to the Bayes error.
   - [x] ... the training error is smaller then the Bayes error.
5. To prevent overfitting one can ...
   - [ ] ... use a smaller test set.
   - [ ] ... use a larger test set.
   - [ ] ... use a smaller training set.
   - [x] ... use a larger training set.
   - [x] ... reduce the capacity of the model.
   - [ ] ... increase the capacity of the model.
6. To prevent underfitting one can ...
   - [ ] ... use a smaller test set.
   - [ ] ... use a larger test set.
   - [ ] ... use a smaller training set.
   - [ ] ... use a larger training set.
   - [ ] ... reduce the capacity of the model.
   - [x] ... increase the capacity of the model.
7. The goal of regularization is to reduce ...
   - [ ] ... the training error.
   - [x] ... the generalization error.
   - [x] ... the test error.
   - [ ] ... the Bayes error.
8. Mark the correct statements and correct the wrong ones.
   - [x] The test set is used to estimate the generalization error.
   - [ ] The ~~training~~ **validation** set is used to control the training.
   - [ ] The ~~validation~~ **training** set is used to learn the task.
   - [ ] The training error typically underestimates the generalization error by a ~~smaller~~ **larger** amount than the validation error.
   - [x] The validation set is used to learn the hyperparameters.

9. Given is a set of samples $\{x^{(1)}..., x^{(m)}\}$ that are independently and identically distributed according to a uniform distribution on the interval $[-0.8,\ 1.2]$.

   (a) Check whether the sample mean

$$\hat{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \tag{1.1}$$

   is an unbiased estimator of the true mean μ.

   (b) Assume that the absolute value of each sample is accidentally taken before the sample mean value is calculated. Thus the new estimator is

$$\hat{\mu}_m = \frac{1}{m}\sum_{i=1}^{m} \left|x^{(i)}\right|. \tag{1.2}$$

   Determine the bias of this poor estimator to the mean μ of the initial distribution.

   (c) How can the estimator of (b) be fixed so that he still gives an unbiased estimate?

Solution:

   (a) The samples are distributed according to the distribution

$$p(x^{(i)}) = \begin{cases} \frac{1}{2}, & -0.8 \le x^{(i)} \le 1.2 \\ 0, & \text{otherwise} \end{cases},$$

   which results in the true mean value of

$$\mu = \frac{a+b}{2} = \frac{(-0.8)+1.2}{2} = 0.2.$$

   The expected value of the sample mean (1.1) is

$$\begin{aligned} \mathbb{E}[\hat{\mu}_m] &= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} x^{(i)}\right] \\ &= \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\left[x^{(i)}\right] \\ &= \frac{1}{m}\sum_{i=1}^{m}\left(\int_{-0.8}^{1.2} x^{(i)}\,\frac{1}{2}\,dx^{(i)}\right) \\ &= \frac{1}{m}\sum_{i=1}^{m} 0.2 \\ &= 0.2, \end{aligned}$$

   resulting in a bias of 0:

$$\begin{aligned} \text{bias}(\hat{\mu}_m) &= \mathbb{E}[\hat{\mu}_m] - \mu \\ &= 0.2 - 0.2 \\ &= 0. \end{aligned}$$

   Thus, the sample mean (1.1) is an unbiased estimator.

(b) To determine the bias of the modified sample mean (1.2), its expected value has to be calculated as follows:

$$
\begin{aligned}
\mathbb{E}[\hat{\mu}_m] &= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\left|x^{(i)}\right|\right] \\
&= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left[\left|x^{(i)}\right|\right] \\
&= \frac{1}{m}\sum_{i=1}^{m}\left(\int_{-0.8}^{1.2}\left|x^{(i)}\right|\,\frac{1}{2}\,dx^{(i)}\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}\left(\int_{-0.8}^{0}\left|x^{(i)}\right|\,\frac{1}{2}\,dx^{(i)} + \int_{0}^{1.2}\left|x^{(i)}\right|\,\frac{1}{2}\,dx^{(i)}\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}\left(\int_{-0.8}^{0} -x^{(i)}\,\frac{1}{2}\,dx^{(i)} + \int_{0}^{1.2} x^{(i)}\,\frac{1}{2}\,dx^{(i)}\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}(0.16 + 0.36) \\
&= \frac{1}{m}\sum_{i=1}^{m}0.52 \\
&= 0.52.
\end{aligned}
$$

This results in a bias of

$$
\begin{aligned}
\operatorname{bias}(\hat{\mu}_m) &= \mathbb{E}[\hat{\mu}_m] - \mu \\
&= 0.52 - 0.2 \\
&= 0.32.
\end{aligned}
$$

Thus, the modified sample mean (1.2) clearly is an biased estimator.

(c) To fix this biased estimator, one can either just subtract the bias term

$$
\begin{aligned}
\hat{\mu}_m &= \frac{1}{m}\sum_{i=1}^{m}\left|x^{(i)}\right| - \operatorname{bias}(\hat{\mu}_m) \\
&= \frac{1}{m}\sum_{i=1}^{m}\left|x^{(i)}\right| - 0.32
\end{aligned}
$$

or add a large number C to each sample, which must be subtracted again after the mean value has been calculated:

$$
\hat{\mu}_m = \frac{1}{m}\sum_{i=1}^{m}\left|x^{(i)} + C\right| - C.
$$

This number must be greater than the absolute value of the lower limit of the interval ($C > |a|$).

The second approach is simpler, since the bias does not have to be calculated.

10. **Optional:** Consider a set of samples $\{x^{(1)}..., x^{(m)}\}$ that are independently and identically distributed according to a uniform distribution on the interval $[0, \theta]$, thus

$$p(x^{(i)}, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x^{(i)} \leq \theta \\ 0, & \text{otherwise} \end{cases}.$$

A biased estimator for the parameter θ is

$$\hat{\theta} = \max(x^{(1)}, \ldots, x^{(m)}).$$

Correct this estimator so that it becomes an unbiased estimator for θ.

Solution:

In order to calculate the expected value $\mathbb{E}\left[\hat{\theta}\right]$, the probability density function must first be calculated from the distribution function of $\hat{\theta}$. The distribution function of X on the interval $[0, \theta]$ is

$$F(x) = P(X \leq x) = \frac{x}{\theta}.$$

Hence, the distribution function of $\hat{\theta} = \max(x^{(1)}, \ldots, x^{(m)})$ is

$$\begin{aligned} P(\hat{\theta} \leq x) &= P(\max(x^{(1)}, \ldots, x^{(m)}) \leq x) \\ &= P(x^{(1)} \leq x \cap \ldots \cap x^{(m)} \leq x) \\ &= P(x^{(1)} \leq x) \cdot \ldots \cdot P(x^{(m)} \leq x) \\ &= \frac{x}{\theta} \cdot \ldots \cdot \frac{x}{\theta} \\ &= \frac{x^m}{\theta^m}. \end{aligned}$$

By derivating this function, the probability density function can be calculated as:

$$p_{\hat{\theta}}(x) = \frac{m\, x^{m-1}}{\theta^m}.$$

With this it is now possible to calculate the expected value

$$\begin{aligned} \mathbb{E}\left[\hat{\theta}\right] &= \int_0^\theta x\, p_{\hat{\theta}}(x)\, dx \\ &= \int_0^\theta x\, \frac{m\, x^{m-1}}{\theta^m}\, dx \\ &= \frac{m}{\theta^m} \int_0^\theta x^m\, dx \\ &= \frac{m}{\theta^m} \left[\frac{x^{m+1}}{m+1}\right]_0^\theta \\ &= \frac{m}{\theta^m} \cdot \frac{\theta^{m+1}}{m+1} \\ &= \frac{m}{m+1}\, \theta. \end{aligned}$$

Except for factor $\frac{m}{m+1}$, this corresponds exactly to the required interval length θ. Hence, an unbiased estimator for the parameter θ is

$$\hat{\theta} = \frac{m+1}{m} \max(x^{(1)}, \ldots, x^{(m)}).$$