Machine Learning A

2021-2022

Home Assignment 2

Yevgeny Seldin **Christian Igel**
Department of Computer Science
University of Copenhagen

The deadline for this assignment is **21 September 2021, 22:00**. You must submit your *individual* solution electronically via the Absalon home page.

A solution consists of:

- A PDF file with detailed answers to the questions, which may include graphs and tables if needed. Do *not* include your source code in the PDF file.
- A .zip file with all your solution source code with comments about the major steps involved in each question (see below). Source code must be submitted in the original file format, not as PDF. The programming language of the course is Python.
- IMPORTANT: Do NOT zip the PDF file, since zipped files cannot be opened in speed grader. Zipped pdf submissions will not be graded.
- Your PDF report should be self-sufficient. I.e., it should be possible to grade it without opening the .zip file. We do not guarantee opening the .zip file when grading.
- Your code should be structured such that there is one main file (or one main file per question) that we can run to reproduce all the results presented in your report. This main file can, if you like, call other files with functions, classes, etc.
- Handwritten solutions will not be accepted, please use the provided latex template to write your report.

1 Illustration of Markov's, Chebyshev's, and Hoeffding's Inequalities (23 points)

2.a Make 1,000,000 repetitions of the experiment of drawing 20 i.i.d. Bernoulli random variables $X_1, \dots, X_{20}$ (20 coins) with bias $\frac{1}{2}$ and answer the following questions.

1. Plot the empirical frequency of observing $\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha$ for $\alpha \in \{0.5, 0.55, 0.6, \dots, 0.95, 1\}$.
2. Explain why the above granularity of α is sufficient. I.e., why, for example, taking $\alpha = 0.51$ will not provide any extra information about the experiment.
3. In the same figure plot the Markov's bound[1] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$.
4. In the same figure plot the Chebyshev's bound[2] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$. (You may have a problem calculating the bound for some values of α. In that case and whenever the bound exceeds 1, replace it with the trivial bound of 1, because we know that probabilities are always bounded by 1.)
5. In the same figure plot the Hoeffding's bound[3] on $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$.
6. Compare the four plots.
7. For $\alpha = 1$ and $\alpha = 0.95$ calculate the exact probability $\mathbb{P}\left(\frac{1}{20}\sum_{i=1}^{20} X_i \geq \alpha\right)$. (No need to add this one to the plot.)

2.b Repeat the question with $X_1, \dots, X_{20}$ with bias 0.1 (i.e., $\mathbb{E}[X_1] = 0.1$) and $\alpha \in \{0.1, 0.15, \dots, 1\}$.

2.c Discuss the results.

Do not forget to put axis labels and a legend in your plot!

---

[1]Markov's bound is the right hand side of Markov's inequality.
[2]Chebyshev's bound is the right hand side of Chebyshev's inequality.
[3]Hoeffding's bound is the right hand side of Hoeffding's inequality.

2 The Role of Independence (13 points)

Design an example of identically distributed, but *dependent* Bernoulli random variables $X_1, \ldots, X_n$ (i.e., $X_i \in \{0, 1\}$), such that

$$\mathbb{P}\left(\left|\mu - \frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \frac{1}{2}\right) = 1,$$

where $\mu = \mathbb{E}[X_i]$.

Note that in this case $\frac{1}{n}\sum_{i=1}^{n} X_i$ does not converge to μ as n goes to infinity. The example shows that independence is crucial for convergence of empirical means to the expected values.

3 Tightness of Markov's Inequality (13 points)

In the previous question you have seen that Markov's inequality may be quite loose. In this question we will show that in some situations it is actually tight. Let ε^* be fixed. Design an example of a random variable X for which

$$\mathbb{P}(X \geq \varepsilon^*) = \frac{\mathbb{E}[X]}{\varepsilon^*}.$$

Prove that the above equality holds for your random variable.

Hint: It is possible to design an example satisfying the above requirement with a random variable X that accepts just two possible values. What should be the values and the probabilities that X accepts these values?

4 The effect of scale (range) and normalization of random variables in Hoeffding's inequality (13 points)

Prove that Corollary 2.5 in Yevgeny's lecture notes (simplified Hoeffding's inequality for random variables in the $[0, 1]$ interval) follows from Theorem 2.3 (general Hoeffding's inequality). [Showing this for one of the two inequalities is sufficient.]

5 Preprocessing & Regularization (38)

Read section 9.1 in e-Chapter 9 of the textbook (Abu-Mostafa et al., 2012). The chapter can be downloaded from `http://book.caltech.edu/bookforum/showthread.php?t=4548`, the login is `bookreaders` and the password the first word on page 27 of the textbook. You can also find a scanned version of the section on Absalon. It is also recommended to read Section 4.2 of the textbook (Abu-Mostafa et al., 2012) on regularization. You can also find a scanned version of the section on Absalon. Note that the *in-sample error* E_{in} corresponds to what we call the empirical risk (or training error).

Solve the first 4 parts (a)–(d) from Exercise 9.4 on page w-Chap:9–4 from the textbook (Abu-Mostafa et al., 2012).

References

Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin. *Learning from Data*. AMLbook, 2012.