Machine Learning B

2022-2023

Home Assignment 1

Yevgeny Seldin **Sadegh Talebi**
Department of Computer Science
University of Copenhagen

The deadline for this assignment is **November 29, 2021, 18:00**. 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 full source code in the PDF file, only selected lines if you are asked to do so.
- 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.
- 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. 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 The Airline Question (40 points)

1. An airline knows that any person making a reservation on a flight will not show up with probability of 0.05 (5 percent). They introduce a policy to sell 100 tickets for a flight that can hold only 99 passengers. Bound the probability that the number of people that show up for a flight will be larger than the number of seats (assuming they show up independently).

2. An airline has collected an i.i.d. sample of 10000 flight reservations and figured out that in this sample 5 percent of passengers who made a reservation did not show up for the flight. They introduce a policy to sell 100 tickets for a flight that can hold only 99 passengers. Bound the probability of observing such sample and getting a flight overbooked.

   There are multiple ways to approach this question. We will guide you through two options. You are asked to solve the question in both ways.

   (a) Let p be the true probability of showing up for a flight (remember that p is unknown). In the first approach we consider two events: the first is that in the sample of 10000 passengers, where each passenger shows up with probability p, we observe 95% of show-ups. The second event is that in the sample of 100 passengers, where each passenger shows up with probability p, everybody shows up. Note that these two events are independent. Bound the probability that they happen simultaneously assuming that p is known. And then find the worst case p (you can do this numerically). With a simple approach you can get a bound of around 0.61. If you are careful and use the right bounds you can get down to around 0.0068.

   It is advised to visualize the problem (the $[0, 1]$ interval with 0.95 point for the 95% show-ups and 1 for the 100% show-ups and p somewhere in $[0, 1]$). This should help you to understand the problem; to understand where the worst case p should be; and to understand in what direction of inequalities you need.

   Attention: This is a frequentist rather than a Bayesian question. In case you are familiar with the Bayesian approach, it cannot be applied here, because we do not provide a prior on p. In case you are unfamiliar with the Bayesian approach, you can safely ignore this comment.

   (b) The second approach considers an alternative way of generating the two samples, using the same idea as in the proof of the VC-bound. Consider the following process of generating the two samples:

      i. We sample 10100 passenger show up events independently at random according to an unknown distribution p.

ii. And then we split them into 10000 passengers in the collected sample and 100 passengers booked for the 99-seats flight.

Bound the probability of observing a sample of 10000 with 95% show ups and a 99-seats flight with all 100 passengers showing up by following the above sampling protocol. If you do things right, you can get a bound of about 0.0062 (there may be some variations depending on how exactly you do the calculation).

2 The Growth Function (10 points)

1. Let $\mathcal{H}$ be a finite hypothesis set with $|\mathcal{H}| = M$ hypotheses. Prove that $m_{\mathcal{H}}(n) \leq \min\{M, 2^n\}$.
2. Let $\mathcal{H}$ be a hypothesis space with 2 hypotheses (i.e., $|\mathcal{H}| = 2$). Prove that $m_{\mathcal{H}}(n) = 2$. (Put attention that you are asked to prove the equality, $m_{\mathcal{H}}(n) = 2$, not an inequality.)
3. Prove that $m_{\mathcal{H}}(2n) \leq m_{\mathcal{H}}(n)^2$.

3 KKT Conditions (20 points)

Consider the following optimization problem:

$$\begin{aligned} \mathsf{P}: \qquad \min \ & \sum_{i=1}^{d} x_i \log\left(\frac{x_i}{a_i}\right) && (1) \\ \text{subject to: } & \sum_{i=1}^{d} x_i \geq 1 && (2) \\ & x_i \in [0,1], \quad \forall i = 1, \ldots, d && (3) \end{aligned}$$

where $a_1, \ldots, a_d$ are some positive constants.

(a) Rewrite P using the standard form (see the definition in the slides or Definition 3.1 in Christian's lecture notes).

(b) Show that P is a convex optimization problem.

(c) Write down the Lagrangian function for P.

(d) Write down the KKT conditions (Sufficiency Theorem) for P.

(e) Form the dual function and dual problem associated to P.

4 Gradient Descent (30 points)

In this exercise, you will implement Gradient Descent (e.g., in Python) and examine its performance on some unconstrained convex problems.

(a) Consider the following optimization problem:

$$\mathsf{P1}: \qquad \min_{x=(x_1,x_2)\in\mathbb{R}^2} f(x_1,x_2) = (x_1+1)^2 + \gamma x_2^2$$

where $\gamma > 0$ is a parameter.

- Show that $\mathsf{P1}$ is a convex optimization problem.
- Solve the problem using Gradient Descent with $\gamma = 2$ using step-size $\alpha_t = \frac{1}{t+1}$. Report the convergence error curves for $f(x_t) - f(x^\star)$ and $\|x_t - x^\star\|_2$.

  Hint: You can easily find $x^\star$ by solving $\nabla f(x_1,x_2) = 0$ by hand.
- Repeat the previous part with $\gamma = 10$ using step-size $\alpha_t = \frac{1}{t+1}$.

(b) Repeat (a) for the following problem:

$$\mathsf{P2}: \qquad \min_{x=(x_1,x_2)\in\mathbb{R}^2} f(x_1,x_2) = e^{x_1+4x_2-0.3} + e^{x_1-4x_2-0.3} + e^{-x_1-0.3}$$

Hint: One can solve $\nabla f(x_1,x_2) = 0$ by hand – which is rather surprising considering the shape of f – and obtain $x^\star = (-\frac{1}{2}\log(2), 0) \approx (-0.346574, 0)$.