Mathematical institute
Prof. Dr. Moritz Diehl

Léo Simpson

Homework accompanying the lecture "Basics in Applied Mathematics"

Homework 12

Hand in: Tuesday, 21.01.2025, after the lecture in the mailbox at the Math Institut
(Don't forget to put your name on your homework.
Please hand in your solutions in groups of two.)

Exercise 1 (Convex sets; 2 points)

Which of the following sets $\mathcal{X} \subset \mathbb{R}^n$ or $\mathcal{A} \subset \mathbb{R}^{n\times n}$ are convex? Justify your answer.

a) $\mathcal{X} = \left\{ x \in \mathbb{R}^n \text{ such that } \sum_{i=1}^{n} |x_i| \le 1 \right\}$

b) $\mathcal{X} = \left\{ x \in \mathbb{R}^3 \text{ such that } x_1 = x_2 \cdot x_3 \right\}$

c) $\mathcal{A} = \left\{ A \in \mathbb{R}^{n\times n} \text{ such that } A = A^\top \right\}$

d) $\mathcal{A} = \left\{ A \in \mathbb{R}^{n\times n} \text{ such that } A = A^\top \text{ and } \forall x \in \mathbb{R}^n,\ x^\top A x \ge 0 \right\}$

Exercise 2 (Convex functions; 2 points)

Which of the following functions $f : \mathbb{R}^2 \to \mathbb{R}$ are convex? Justify your answer.

a) $f(x, y) = xy$

b) $f(x, y) = e^{2x-3y} + 4y$

c) $f(x, y) = \sin(x) + \cos(y)$

d) $f(x, y) = \max\{x, y\} = \begin{cases} x & \text{if } x \ge y \\ y & \text{if } y > x \end{cases}$

Exercise 3 (Jensen Inequality ; 3 points)

a) Let $\mathcal{X} \subset \mathbb{R}^n$ be a convex set. Let $x_1, \dots, x_m$ be some elements of $\mathcal{X}$. Show that the average point $\frac{x_1+\cdots+x_m}{m}$ is also an element of $\mathcal{X}$.

<u>Hint 1:</u> prove this property via induction.

<u>Hint 2:</u> find some α such that: $\frac{x_1+\cdots+x_{m+1}}{m+1} = (1-\alpha)\frac{x_1+\cdots+x_m}{m} + \alpha x_{m+1}$.

b) Now let $f : \mathcal{X} \to \mathbb{R}$ be a convex function. Show the following inequality:

$$f\left(\frac{x_1 + \cdots + x_m}{m}\right) \le \frac{f(x_1) + \cdots + f(x_m)}{m} \tag{1}$$

c) Now show the following generalization:

$$f\left(\sum_{j=1}^{m} \alpha_j x_j\right) \le \sum_{j=1}^{m} \alpha_j f(x_j) \tag{2}$$

for any $\alpha_1, \dots, \alpha_m \ge 0$ such that $\sum_{j=1}^{m} \alpha_j = 1$.

Exercise 4 (Minimizer of the Cross Entropy; 2 points)

Let $\mathcal{P}_m$ be the set of probability distributions over the set $\{1, \ldots, m\}$, i.e.:

$$\mathcal{P}_m = \left\{ p \in \mathbb{R}^m \text{ such that } \forall j,\ p_j \geq 0, \text{ and } \sum_{j=1}^{m} p_j = 1 \right\} \tag{3}$$

a) Prove that for any $p, q \in \mathcal{P}_m$, the inequality holds:

$$\sum_{j=1}^{m} p_j \log\left(\frac{q_j}{p_j}\right) \leq 0 \tag{4}$$

<u>Hint:</u> Apply the Jensen inequality (2) to the function $f(x) = -\log(x)$ (after proving that this is a convex function).

b) We define the cross entropy between two distributions as follows:

$$L(p, q) = -\sum_{j=1}^{m} p_j \log(q_j) \tag{5}$$

Let p be an element of $\mathcal{P}_m$. Show that $q = p$ is a minimizer of:

$$\underset{q \in \mathcal{P}_m}{\text{minimize}}\ L(p, q) \tag{6}$$

<u>Note:</u> You do not have to show that it is the unique minimizer.

Exercise 5 (Mean and variance estimation; 5 points)

In this part, we consider a regression task where the dataset takes the following form: $(a_1, y_1), \ldots, (a_m, y_m)$ with $a_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$ (single input and single ouput problem).
The goal is to identify a model of the form:

$$y_j \approx \theta a_j + b$$

Furthermore, we also want to estimate the standard deviation of the model, i.e. the scale σ of the errors $y_j - (\theta a_j + b)$.
For that purpose, we propose the following problem:

$$\underset{\theta, b, \sigma > 0}{\text{minimize}}\ \underbrace{\frac{1}{m} \sum_{j=1}^{m} \frac{(y_j - (\theta a_j + b))^2}{\sigma} + \sigma}_{=: f(\theta, b, \sigma)}$$

We will define the parameter vector $x := (\theta, b, \sigma) \in \mathbb{R}^3$, and define the feasible set $\mathcal{X}$ as follows:

$$\mathcal{X} = \left\{ x = (\theta, b, \sigma) \in \mathbb{R}^3 \text{ such that } \sigma > 0 \right\}$$

a) Prove that the set $\mathcal{X}$ is convex.

b) Let us define the function $g : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}$ defined as:

$$g(e, \sigma) = \frac{e^2}{\sigma} + \sigma$$

Prove that the function g is convex.

c) Use the previous question to show that the function $f(\cdot, \cdot, \cdot)$ is convex over $\mathcal{X}$.

d) Using the previous questions and results from the lectures, write down the equations that are necessary and sufficient for a point $x^\star = (\theta^\star, b^\star, \sigma^\star)$ to be a minimizer of the problem (these equations should be explicit for this problem, *not* the generic equation from the lecture).

e) Express the solution $x^\star = (\theta^\star, b^\star, \sigma^\star)$ explicitly.

Note: You might want to simplify the expressions by defining the following variables:

$$\bar{y} := \frac{1}{m} \sum_{j=1}^{m} y_j \qquad \bar{a} := \frac{1}{m} \sum_{j=1}^{m} a_j$$

$$\bar{c} := \frac{1}{m} \sum_{j=1}^{m} a_j y_j \qquad \bar{h} := \frac{1}{m} \sum_{j=1}^{m} a_j^2$$

Note: You can first express $\theta^\star$ in terms of the other variables, then $b^\star$ as a function of $\theta^\star$, and then express $\sigma^\star$ as a function of $\theta^\star$ and $b^\star$.

Exercise 6 (Programming exercise; 4 points)

Open the jupyter notebook, and fill in the missing parts of the code.