Assignment 8

Submission Deadline: **18 November, 2025** at 23:59
Course Website: https://ti.inf.ethz.ch/ew/courses/LA25/index.html

Exercises

You can get feedback from your TA for Exercise 2 by handing in your solution as pdf via Moodle before the deadline.

1. Linear regression (in-class) (★☆☆)

We want to determine the parameters of a certain model function from some measured values. Assume that we measured the following values

t_i	1	2	3	4	5
b_i	2	3	5	6	8

where $i \in [5]$. Moreover, assume that we want to model the relationship between t, b by a function f, i.e. $b = f(t)$. In this exercise, we restrict f to be a line, i.e. f should have the form

$$f(t) = \alpha_1 t + \alpha_0$$

for parameters $\alpha_1, \alpha_0 \in \mathbb{R}$. Our goal is to find suitable values for α_1, α_0. As discussed in the lecture, this idea of fitting a line through a set of datapoints is called linear regression.

a) For each datapoint (t_i, b_i) with $i \in [5]$, we get an equation for α_1, α_0 from $f(t_i) = b_i$. Write down the system of linear equations that we get by combining all five equations.

b) Do you expect this system to have any solutions? (Answer this intuitively without actually solving the system).

c) Using the normal equations, find an approximate solution to the system you wrote down.

2. Fitting a parabola (hand-in) (★★☆)

We want to determine the parameters of a certain model function from some measured values. Assume that we measured the following values

t_i	-2	-1	0	1	2
b_i	3	2	1	4	5

where $i \in [5]$. Moreover, assume that we want to model the relationship between t, b by a function f, i.e. $b = f(t)$. In this exercise, we restrict f to be a parabola, i.e. f should have the form

$$f(t) = \alpha_2 t^2 + \alpha_1 t + \alpha_0$$

for parameters $\alpha_2, \alpha_1, \alpha_0 \in \mathbb{R}$. Our goal is to find suitable values for $\alpha_2, \alpha_1, \alpha_0$.

a) We need the following extension of Lemma 6.1.2 in the lecture notes: Let $m \in \mathbb{N}_{\geq 3}$ and $t_1, \ldots, t_m \in \mathbb{R}$ with $t_i - t_j \neq 0$ for all different $i, j \in [m]$. Consider

$$A = \begin{bmatrix} t_1^2 & t_1 & 1 \\ \vdots & \vdots & \vdots \\ t_m^2 & t_m & 1 \end{bmatrix}.$$

Show that A has full column rank. Does the converse hold, that is, suppose that A has full column rank, do we have $t_1, \ldots, t_m \in \mathbb{R}$ with $t_i - t_j \neq 0$ for all different $i, j \in [m]$?

b) Following Exercise 1, write down the system of linear equations

$$A \begin{pmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{pmatrix} = b$$

that you obtain from $f(t_i) = b_i$. Compute suitable values for $\alpha_2, \alpha_1, \alpha_0$ that minimize

$$\left\| A \begin{pmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_0 \end{pmatrix} - b \right\|^2.$$

Is the solution unique?

3. Alternative definition of a projection (★☆☆)

Let $S \subseteq \mathbb{R}^m$ be a subspace. From the lecture, we know that for every $\mathbf{v} \in \mathbb{R}^m$ there exist unique vectors $\mathbf{s} \in S$ and $\mathbf{t} \in S^\perp$ such that

$$\mathbf{v} = \mathbf{s} + \mathbf{t}$$

We can then define the projection onto S as the function $P_S : \mathbb{R}^m \to \mathbb{R}^m$ given by

$$P_S(\mathbf{v}) = \mathbf{s}$$

Using this equivalent definition, answer the following questions:

a) Prove that the function P_S is linear.

b) Prove that $P_S^2 = P_S$ (applying the projection twice has the same effect as applying it once).

c) Determine the kernel and image of P_S.

d) Show that for every $\mathbf{v} \in \mathbb{R}^m$, the following identities hold:

$$(P_S + P_{S^\perp})(\mathbf{v}) = \mathbf{v}, \qquad (P_S \circ P_{S^\perp})(\mathbf{v}) = \mathbf{0}$$

4. Subspaces with intersection $\{0\}$ (★★☆)

Let U, W be subspaces of a vector space V satisfying $U \cap W = \{\mathbf{0}\}$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a set of n linearly independent vectors in U, and let $\mathbf{w}_1, \ldots, \mathbf{w}_m$ be a set of m linearly independent vectors in W. Prove that the set of $n + m$ vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{w}_1, \ldots, \mathbf{w}_m$ is linearly independent in V.

Hint: *Note that this statement is formulated for vector spaces in general. If you are unsure with arguing about arbitrary vector spaces, consider first proving it for the special case $V = \mathbb{R}^k$ for some $k \geq m + n$.*

5. Weighted linear regression (★★☆)

Assume we are given $m \in \mathbb{N}^+$ datapoints $(t_1, b_1), \dots, (t_m, b_m)$ where $t_k, b_k \in \mathbb{R}$ for all $k \in [m]$. For convenience, we define

$$A := \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \text{ and } \mathbf{b} := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Assume that we are additionally given $\lambda_1, \dots, \lambda_m \in \mathbb{R}^+$ (i.e. $\lambda_i > 0$ for all $i \in [m]$), and let $\Lambda \in \mathbb{R}^{m \times m}$ denote the diagonal matrix with $\lambda_1, \dots, \lambda_m$ on its diagonal. The goal of this exercise is to find α_0 and α_1 in $\mathbb{R}$ such that

$$\sum_{k=1}^{m} \lambda_k (b_k - (\alpha_0 + \alpha_1 t_k))^2$$

is minimized. Intuitively speaking, the weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}^+$ are used to put an emphasize on some datapoints that we think of as more important. For example, if λ_i is large for some $i \in [m]$ (relative to the other weights), this means that is it relatively important for us to get close to this datapoint.

a) Prove that $A^\top \Lambda A$ is invertible if and only if $A^\top A$ is invertible.

b) Suppose that A has full column rank. Prove that the optimal choice for α_0 and α_1 is given by

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = (A^\top \Lambda A)^{-1} A^\top \Lambda \mathbf{b}.$$

6. Fitting a line (★★★)

Assume we are given $m \geq 2$ distinct datapoints $(t_1, b_1), \dots, (t_m, b_m)$ where $t_k, b_k \in \mathbb{R}$ for all $k \in [m]$ (distinct means that we have $t_i \neq t_j$ for all $i \neq j$ with $i, j \in [m]$). Using the least squares method, we want to find a line described by two parameters $\alpha_0, \alpha_1 \in \mathbb{R}$ such that we have

$$b_k \approx \alpha_0 + \alpha_1 t_k$$

for all $k \in [m]$. More concretely, we want to solve the optimization problem

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^2} ||A\boldsymbol{\alpha} - \mathbf{b}||^2 = \min_{\alpha_0, \alpha_1 \in \mathbb{R}} \sum_{k=1}^{m} (b_k - (\alpha_0 + \alpha_1 t_k))^2$$

where

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}.$$

Remark 6.1.3 in the lecture notes gives the closed form solution

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{m} \sum_{k=1}^{m} b_k \\ (\sum_{k=1}^{m} t_k b_k)/(\sum_{k=1}^{m} t_k^2) \end{pmatrix}$$

for this problem under the additional assumption that $\sum_{k=1}^{m} t_k = 0$. In this exercise, we want to use this to find a closed form solution for the general case, i.e. we want to drop the assumption $\sum_{k=1}^{m} t_k = 0$.

a) Let $c \in \mathbb{R}$ be some constant and consider new datapoints $(t'_1, b_1), \ldots, (t'_m, b_m)$ with $t'_k = t_k + c$ for all $k \in [m]$. This gives us a new optimization problem

$$\min_{\boldsymbol{\alpha}' \in \mathbb{R}^2} ||A'\boldsymbol{\alpha}' - \mathbf{b}||^2 = \min_{\alpha'_0, \alpha'_1 \in \mathbb{R}} \sum_{k=1}^{m} (b_k - (\alpha'_0 + \alpha'_1 t'_k))^2$$

where

$$\boldsymbol{\alpha}' = \begin{pmatrix} \alpha'_0 \\ \alpha'_1 \end{pmatrix}, \quad A' = \begin{bmatrix} 1 & t'_1 \\ \vdots & \vdots \\ 1 & t'_m \end{bmatrix}.$$

Intuitively speaking, how do the optimal solutions $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ of the two optimization problems compare? Do we expect to have $\alpha_0 = \alpha'_0$? Do we expect to have $\alpha_1 = \alpha'_1$? Give a brief intuitive argument.

b) As discussed in the lecture notes, we want to set $c = -\frac{1}{m}\sum_{k=1}^{m} t_k$ so that the columns of A' will be orthogonal. Verify that this is indeed the case, i.e. verify that the columns of A' defined as above with $c = -\frac{1}{m}\sum_{k=1}^{m} t_k$ are orthogonal.

c) Given $\boldsymbol{\alpha}'$ such that $||A'\boldsymbol{\alpha}' - \mathbf{b}||^2$ is minimized (i.e. $\boldsymbol{\alpha}'$ is an optimal solution), prove that

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \begin{pmatrix} c\alpha'_1 \\ 0 \end{pmatrix}$$

minimizes $||A\boldsymbol{\alpha} - \mathbf{b}||^2$ (i.e. $\boldsymbol{\alpha}$ is an optimal solution for the original problem).

***Hint**: This subtask gives away the answer to a), but make sure that you have some intuition of why we expect $\alpha'_1 = \alpha_1$.*

d) Note that by subtask b), we can use the closed form solution from Remark 6.1.3 to solve

$$\min_{\boldsymbol{\alpha}' \in \mathbb{R}^2} ||A'\boldsymbol{\alpha}' - \mathbf{b}||^2.$$

Combine this with subtask c) to get a closed form solution for the original problem

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^2} ||A\boldsymbol{\alpha} - \mathbf{b}||^2 = \min_{\alpha_0, \alpha_1 \in \mathbb{R}} \sum_{k=1}^{m} (b_k - (\alpha_0 + \alpha_1 t_k))^2.$$

You do not need to simplify the formula you get.