Afonso Bandeira
Bernd Gärtner

Assignment 8

Submission Deadline: **21 November, 2023** at 23:59
Course Website: https://ti.inf.ethz.ch/ew/courses/LA23

Exercises

You can get feedback from your TA for Exercise 1 by handing in your solution as pdf via Moodle before the deadline.

1. **Gram-Schmidt (hand-in)** (★☆☆)

   This task includes Challenge 20 from the lecture notes.

   Consider the invertible matrices

   $$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

   and

   $$B = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & 5 & 6 \\ 0 & 0 & 7 & 8 \\ 0 & 0 & 0 & 9 \end{bmatrix}.$$

   a) Apply the Gram-Schmidt process to the columns of A.

   b) Write down a QR-decomposition of A.

   c) Apply the Gram-Schmidt process to the columns of B.

   d) Is it always true that the Gram-Schmidt process on the columns of an upper triangular $n \times n$ matrix with non-zero diagonal entries yields the canonical basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$? Provide a proof or counterexample.

2. **Permutation matrices** (★☆☆)

   This task includes Challenge 17 from the lecture notes.

   Let $P \in \mathbb{R}^{n \times n}$ be a permutation matrix for some $n \geq 1$. In particular, P has the form

   $$P = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{e}_{p(1)} & \mathbf{e}_{p(2)} & \cdots & \mathbf{e}_{p(n)} \\ | & | & \cdots & | \end{bmatrix}$$

   where $p : [n] \to [n]$ is a bijective function (the permutations of $[n]$ are exactly the bijective funtions $p : [n] \to [n]$). Prove that P is orthogonal.

3. **Orthogonal matrices** (★☆☆)

   This task includes Challenge 16 from the lecture notes.

a) Let R_θ be a 2×2 rotation matrix. Prove that R_θ is orthogonal.

Hint: It might be worth to have another look at the exercise on rotation matrices from Assignment 2.

b) Find an orthogonal 2×2 matrix that is not a rotation matrix.

c) Consider an arbitrary 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Prove that if A is orthogonal, then we have $|ad - bc| = 1$.

d) Prove that the converse is not true, i.e. find values for a, b, c, d such that A is not orthogonal but we still have $|ad - bc| = 1$.

4. Fitting a line (★★☆)

This task includes Challenge 11 from the lecture notes.

As in Section 4.3.2, 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

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

In Remark 4.3.5, we derived the closed form solution

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

for this problem under the additional assumption that $\sum_{k=1}^{m} t_k = 0$. In this exercise, we want 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 $(t'_1, b_1), \dots, (t'_m, b_m)$ 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

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

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

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \begin{bmatrix} c\alpha_1' \\ 0 \end{bmatrix}$$

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 4.3.5 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.

5. Fitting a parabola (★☆☆)

This task includes Challenge 12 from the lecture notes of the second part of the course.

Assume we are given $m \geq 3$ distinct datapoints $(t_1, b_1), \ldots, (t_m, b_m)$ 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 parabola described by three parameters $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$ such that we have

$$b_k \approx \alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2$$

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

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^3} ||A\boldsymbol{\alpha} - \mathbf{b}||^2 = \min_{\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}} \sum_{k=1}^{m} (b_k - (\alpha_0 + \alpha_1 t_k + \alpha_2 t_k^2))^2$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}.$$

a) Compute the matrix $A^\top A$.

b) Prove that for $A^\top A$ to be diagonal, we must have $t_k = 0$ for all $k \in [m]$. Note that this is again an uninteresting case which is actually excluded by the assumption $m \geq 3$ and the assumption that our datapoints are distinct.