Home Assignment 1

Ilana Pervoi, Pan Eyal

1. **Minimizations with different norms lead to different answer:**

a)

$$l_2: \min_c\{(c-x_1)^2+(c-x_2)^2+(c-x_3)^2\} =$$
$$\min_c\{c^2-2cx_1+x_1^2+c^2-2cx_2+x_2^2+c^2-2cx_3+x_3^2\} =$$
$$\min_c\{3c^2-2c(x_1+x_2+x_3)+x_1^2+x_2^2+x_3^2\}$$

We will mark

$$f(c)=3c^2-2c(x_1+x_2+x_3)+x_1^2+x_2^2+x_3^2$$
$$f'(c)=6c-2(x_1+x_2+x_3)$$

By setting it to be 0, we will find the critical point c (which is a minimum) as required:

$$6c-2(x_1+x_2+x_3)=0$$
$$\boldsymbol{c=\frac{1}{3}(x_1+x_2+x_3)}$$

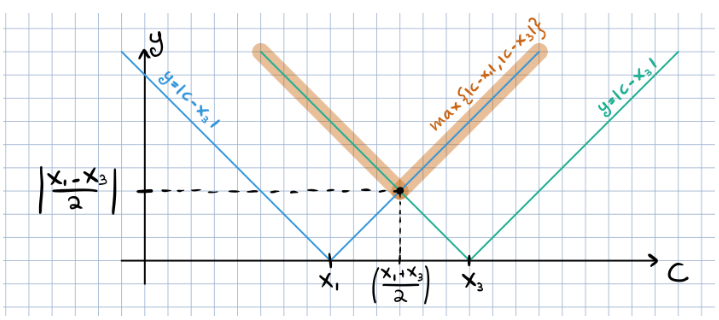
b) $l_\infty: \min_c\{\max\{|c-x_1|,|c-x_2|,|c-x_3|\}\}$

Because $x_1<x_2<x_3$ we can conclude that:

$$\max\{|c-x_1|,|c-x_2|,|c-x_3|\}=\max\{|c-x_1|,|c-x_3|\}$$

we will prove that:

$$\min\{\max\{|c-x_1|,|c-x_3|\}\}=\left|\frac{x_1+x_3}{2}\right|:$$



i. If $c < \left|\frac{x_1+x_3}{2}\right|$ then:

$$|c - x_3| > \left|\frac{x_1 + x_3}{2} - x_3\right| = \left|\frac{x_1 + x_3 - 2x_3}{2}\right| = \left|\frac{x_1 - x_3}{2}\right|$$

So:

$$\min\{\max\{|c - x_1|, |c - x_3|\}\} > \left|\frac{x_1 - x_3}{2}\right|$$

ii. If $c = \left|\frac{x_1+x_3}{2}\right|$ then:

$$\max\{|c - x_1|, |c - x_3|\} =$$
$$\max\left\{\left|\frac{x_1 + x_3}{2} - x_1\right|, \left|\frac{x_1 + x_3}{2} - x_3\right|\right\}$$
$$= \max\left\{\left|\frac{x_3 - x_1}{2}\right|, \left|\frac{x_1 - x_3}{2}\right|\right\} = \left|\frac{x_1 - x_3}{2}\right|$$

iii. If $c > \left|\frac{x_1+x_3}{2}\right|$ then:

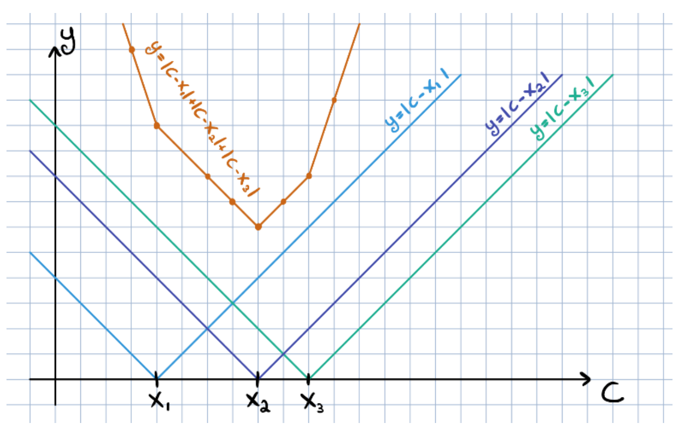
$$|c - x_1| > \left|\frac{x_1 + x_3}{2} - x_1\right| = \left|\frac{x_1 + x_3 - 2x_1}{2}\right| = \left|\frac{x_3 - x_1}{2}\right| = \left|\frac{x_1 - x_3}{2}\right|$$

So:

$$\min\{\max\{|c - x_1|, |c - x_3|\}\} > \left|\frac{x_1 - x_3}{2}\right|$$

Therefore: $\boldsymbol{c} = \left|\frac{\boldsymbol{x_1+x_3}}{\boldsymbol{2}}\right|$ will set the minimal value for $\max\{|c - x_1|, |c - x_3|\}$ ■

c) $l_1: \min_c\{|c - x_1| + |c - x_2| + |c - x_3|\}$



we will prove that $\min_{c}\{|c - x_1| + |c - x_2| + |c - x_3|\} = x_2$:

i. If $c < x_1$ then

$$|c - x_3| > |x_3 - x_1|$$

So:

$$|c - x_1| + |c - x_2| + |c - x_3| > |x_3 - x_1|$$

ii. If $x_1 \leq c < x_2$ then

$$|c - x_3| + |c - x_2| > |x_3 - x_1|$$

So:

$$|c - x_1| + |c - x_2| + |c - x_3| > |x_3 - x_1|$$

iii. If $c = x_2$ then

$$|c - x_1| + |c - x_2| + |c - x_3| = |c - x_1| + 0 + |c - x_3| = |x_3 - x_1|$$

iv. If $x_2 < c \leq x_2$ then

$$|c - x_1| + |c - x_2| > |x_1 - x_3|$$

So:

$$|c - x_1| + |c - x_2| + |c - x_3| > |x_1 - x_3| = |x_3 - x_1|$$

v. If $c > x_3$ then

$$|c - x_1| > |x_1 - x_3|$$

So:

$$|c - x_1| + |c - x_2| + |c - x_3| > |x_3 - x_1|$$

Therefore: $\boldsymbol{c = x_2}$ will set the minimal value for $|c - x_1| + |c - x_2| + |c - x_3|$ ■

2. <u>Least Squares</u>:

a) We compute the code in python using NumPy package.

By calculating the solution x by the normal equation:

$$A^T Ax = A^T b$$

we receive

$$\boldsymbol{x = [\frac{17}{10}, \frac{3}{15}, \frac{7}{10}]}$$

b) No, we won't receive a unique solution.

$A^T A$ is singular, hence as we've learned we will receive infinitely many solutions.

Furthermore, we can see that for each $v \in \ker(A^T A)$ ($|\ker (A^T A)| > \{0\}$ from the singularity properties), $x_0 = x + v$,when x is the solution we already found, is also a valid LS solution.

proof:

$$A^T A x_0 = A^T b$$
$$A^T A(x+n) = A^T b$$
$$A^T(Ax + An) = A^T b$$
$$A^T A x + A^T A n = A^T b$$
(And because $v \in \ker\,(A^T A)$:)
$$A^T A x + 0 = A^T b$$
$$A^T A x = A^T b \ \blacksquare$$

To find the minimal loss, we compute the code by calculating the following r:

$$r = Ax - b$$

Then, by calculating its norm we determine its size:

$$\boldsymbol{|r| = 1.8830013538332964}$$

c) For the 1st equation to almost get exactly satisfied (when $|r_1| < 10^{-3}$),

we've made a loop that will change the weight of the 1st equation by adding to the LS equation a "weighted matrix" we called W that has bigger weight at index [0,0].

Meaning, W is in the form of:

$$\begin{pmatrix} w & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

To find the minimal w value to satisfy the equation, and that $|r_1| < 10^{-3}$, we've made a while loop that increases w value by 0.1 with each iteration, and checks for r_1 loss.

We've exited the loop when the loss was less then 10^{-3}.

We found that it occurred when

$$\boldsymbol{w = 807.3499999994108 \approx 807.35}$$

3. Eigenvalues and positive definite matrices:

a) Let λ be the eigenvalue of matric C.

By definition,

$$Cv = \lambda v$$

$$v - Cv = v - \lambda v$$

$$v = Iv \quad \rightarrow \quad Iv - Cv = v - \lambda v$$

$$(I - C)v = (1 - \lambda)v$$

Which means $(1 - \lambda)$ is an eigenvalue of the matrix $(I - C)$ as required. ■

b) Let $A \in \mathbb{R}^{m \times n}, m \geq n$.

Support claim:

$$Rank(A) = Rank(A^T A)$$

Proof:

($\subseteq$): Let $x \in \ker(A)$

$$Ax = 0$$

$$A^T Ax = A^T 0 = 0$$

$$x \in \ker(A^T A)$$

$$\ker(A) \subseteq \ker(A^T A)$$

($\supseteq$): Let $x \in \ker(A^T A)$

$$A^T Ax = 0$$

$$x^T A^T Ax = x^T 0 = 0$$

$$(Ax)^T (Ax) = 0$$

$$||Ax||^2 = 0$$

By norm properties

$$Ax = 0$$

$$x \in \ker(A)$$

$$\ker(A^T A) \subseteq \ker(A)$$

$$\ker(A) \subseteq \ker(A^T A) \; and \; \ker(A^T A) \subseteq \ker(A) \; \rightarrow \; \ker(A^T A) = \ker(A)$$

And so

$$Rank(A) = Rank(A^T A) \blacksquare$$

Main claim:

(→): Assume A is a full rank.

$$Rank(A) = n$$

By support claim

$$Rank(A^T A) = Rank(A) = n$$

$$A^T A \in \mathbb{R}^{m \times n}$$

$$A^T A \text{ } is \text{ } full \text{ } rank$$

$$A^T A \text{ } is \text{ } invertible$$

(←): Assume $A^T A$ is invertible.

$$A^T A \text{ } is \text{ } full \text{ } rank$$

By support claim

$$n = Rank(A^T A) = Rank(A)$$

$$A \text{ } is \text{ } full \text{ } rank \blacksquare$$

c) We saw in the previous question that $A \text{ } is \text{ } full \text{ } rank \text{ } iff \text{ } A^T A \text{ } is \text{ } invertible.$

By proving that $A^T A \text{ } is \text{ } invertible \text{ } iff \text{ } A^T A \text{ } is \text{ } symmetric \text{ } and \text{ } positive \text{ } definite$ and transitivity properties, we will conclude that

$A \text{ } is \text{ } full \text{ } rank \text{ } iff \text{ } A^T A \text{ } is \text{ } symmetric \text{ } and \text{ } positive \text{ } definite.$

Firstly, $A^T A \text{ } is \text{ } always \text{ } symmetric$ because its transpose is also $A^T A$:

$$(A^T A)^T = A^T (A^T)^T = A^T A \blacksquare$$

(→): Assume $A^T A$ is invertible.

To show that $A^T A \text{ } is \text{ } positive \text{ } definite$ we need to prove that $for \text{ } each \text{ } x \neq 0, x^T A^T Ax > 0.$

$$x^T A^T Ax = (Ax)^T Ax =< Ax, Ax \geq ||Ax||$$

From norm properties

$$||Ax|| \geq 0$$

$$\rightarrow x^T A^T Ax \geq 0$$

Finally, $x^T A^T Ax \neq 0$ because $x \neq 0$: $x^T \neq 0$ and $A^T A \text{ } is \text{ } invertible$ so $A^T Ax \neq 0$

Hence

$$x^T A^T Ax > 0$$

($\leftarrow$): Assume $A^T A$ is positive definite.

By definition,

$$for\ each\ x \neq 0, x^T A^T A x > 0$$

Let $x \neq 0$, then

$$A^T A x \neq 0$$

$$\ker(A^T A) = \{0\}$$

$$Rank(A^T A) = n$$

$$A^T A\ is\ invertible\ \blacksquare$$

d) Assume $\alpha > 0$.

We need to prove that $A^T A + \alpha I$ is always positive definite.

Once again, we'll do so by showing that $for\ each\ x \neq 0,\ \ x^T(A^T A + \alpha I)x > 0$.

Let $x \neq 0$.

$$x^T(A^T A + \alpha I)x = (x^T A^T A + x^T \alpha)x = x^T A^T A x + x^T \alpha x$$

We will notice that

$$x^T A^T A x = (Ax)^T A x = \|Ax\|_2^2$$

And by norm properties

$$x^T A^T A x = \|Ax\|_2^2 \geq 0$$

Also,

$$x^T \alpha x = \alpha x^T x = \alpha \|x\|_2^2$$

$\alpha > 0\ and\ x \neq 0$, therefore:

$$\alpha \|x\|_2^2 > 0$$

In conclusion,

$$x^T(A^T A + \alpha I)x = \underbrace{x^T A^T A x}_{\geq 0} + \underbrace{x^T \alpha x}_{>0} > 0\ \ \blacksquare$$

4. **Frobenius Norm:**

a) We need to prove that $trace(A^T B) =< A[:], B[:] >$

Proof:

$$trace(A^T B) = \sum_{i=1}^{n} (A^T B)_{ii}$$

By matrix multiplication definition: $(A^T B)_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$:

For $(A^T B)_{ii} = \sum_{k=1}^{m} a_{ik} b_{ki}$

$$trace(A^T B) = \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} b_{ki}$$

For: $A[:] = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \\ a_{12} \\ \vdots \\ a_{m2} \\ \vdots \\ \vdots \\ a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}$, $B[:] = \begin{bmatrix} b_{11} \\ \vdots \\ b_{m1} \\ b_{12} \\ \vdots \\ b_{m2} \\ \vdots \\ \vdots \\ b_{1n} \\ \vdots \\ b_{mn} \end{bmatrix}$

$$< A[:], B[:] > =$$

$$= a_{11} b_{11} + \cdots + a_{m1} b_{m1} + a_{12} b_{12} + \cdots + a_{m2} b_{m2} + \cdots \cdots + a_{1n} b_{1n} + \cdots + a_{mn} b_{mn} =$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} b_{ki}$$

Therefore: $trace(A^T B) = \sum_{i=1}^{n} \sum_{k=1}^{m} a_{ik} b_{ki} =< A[:], B[:] >$ ■

b) We want to find $\hat{x}$ such that

$$\hat{x} = argmin\left||AX - B|\right|_F^2$$

$$f(X) = \left||AX - B|\right|_F^2 = trace\left((AX - B)^T (AX - B)\right)$$

And from the previous proof

$$trace\left((AX - B)^T (AX - B)\right) =< (AX - B)[:], (AX - B)[:] >$$

Hence

$$f(X) =< (AX - B)[:], (AX - B)[:] > = \left((AX - B)[:]\right)^T (AX - B)[:]$$

$$= \left(\left((AX)[:]\right)^T - (B[:])^T\right) (AX - B)[:]$$

$$= \left((AX)[:]\right)^T (AX)[:] - \left((AX)[:]\right)^T B[:] - (B[:])^T (AX)[:] + (B[:])^T B[:]$$

We will perform derivative and find minimal value:

$$f'(X[:]) = (A^T AX)[:] + ((A^T A)^T X)[:] - 2((B^T A)^T)[:] + 0 = 2(A^T AX)[:] - 2(A^T B)[:]$$

$$(2A^T AX)[:] - (2A^T B[:]) = 0$$

$$(A^T AX)[:] = (A^T B)[:]$$

In other words,

$$A^T AX = A^T B$$

The solution X for the equation will set the minimal value for $||AX - B||_F^2$.

The solution will be unique when $A^T A$ is invertible.

c) To find the solution D for the matrices A, B we will find, for each $1 \leq i \leq 4$ row of A and B, a LS solution D_{ii} for the equation : $||a_i D_{ii} - b_i||_2^2$.

By calculating, we found out that

$$\boldsymbol{D} = \begin{bmatrix} \mathbf{0.11385057} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1.30588235} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0.25647059} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{6.88571429} \end{bmatrix}$$

5. **Working with real Data:**

At the beginning, we choose all the following countries to work with:

Belgium, Germany, Portugal, Spain, Switzerland, Italy and United Kingdom,

We choose to work with 80% of the data to create our model, by taking the first 189 days.

We defined $A_{base-countries-data}$ to be the matrix that represent the 1st 80% of those countries data in the following way:

- each column of the matrix will represent a country.
- each row will represent the new cases of covid-19 divided by the density of that country.

We defined $B_{base-france-data}$ to be a vector that represents France's data similarly.

Then, we calculated the LS solution x and for the equation:

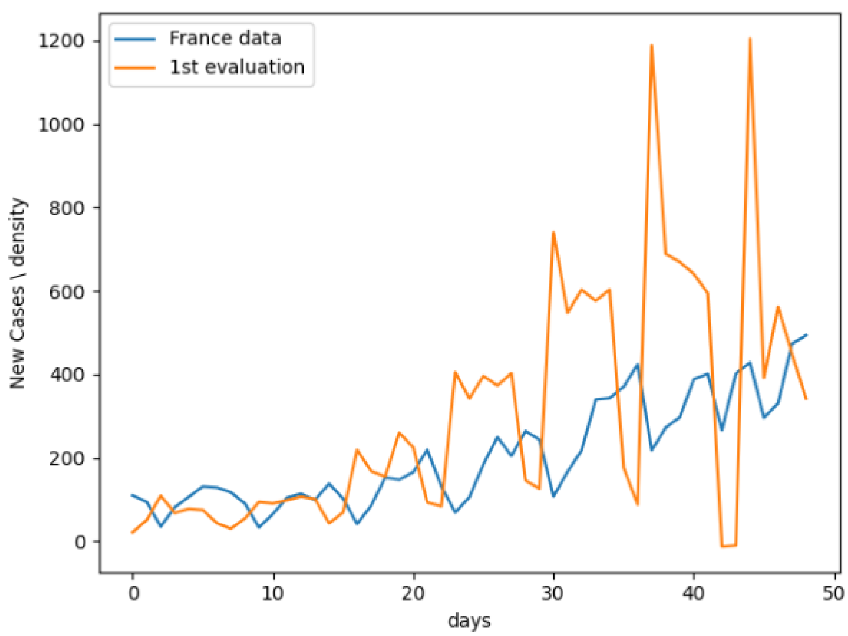
$$A_{base-countries-data} \cdot x = B_{base-france-data}$$

We got that $x = \begin{bmatrix} 8.38437172 \\ -2.17436505 \\ -0.79235153 \\ 1.748401 \\ 0.11856301 \\ 10.1016209 \\ 0.46186982 \end{bmatrix}$ with a minimal error of: 34682.606

We defined $A_{validation-countries-data}$ to be the remaining 20% and then applied our model on it. So:

$$B_{evaluation-france-data} = A_{validation-countries-data} \cdot x$$

We received the following graph:

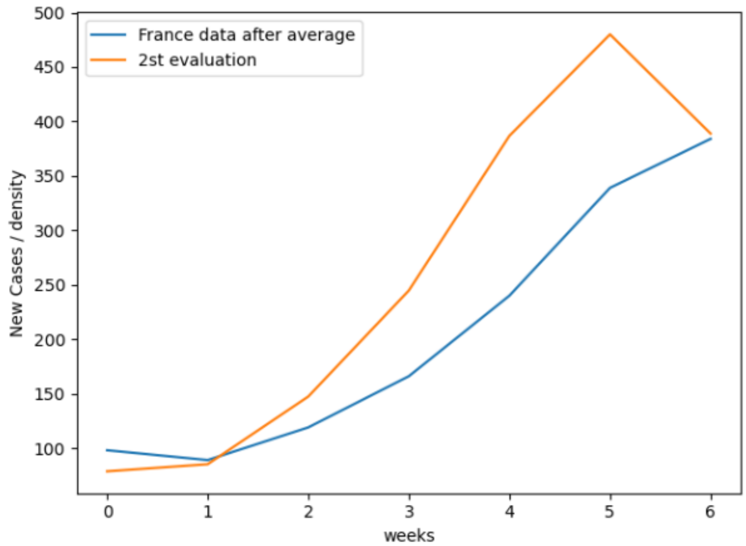


We can see that the graph has a lot of noise, therefore we chose to average it by weeks and normalize it by density of each country.

By repeating the process explained above, with the new reformed data

We got that $x = \begin{bmatrix} 4.9914889 \\ -3.03857838 \\ -0.24425961 \\ 1.7437747 \\ 0.41726226 \\ 8.69233556 \\ 0.7205325 \end{bmatrix}$ with a minimal error of: 185.727

We received the following graph:

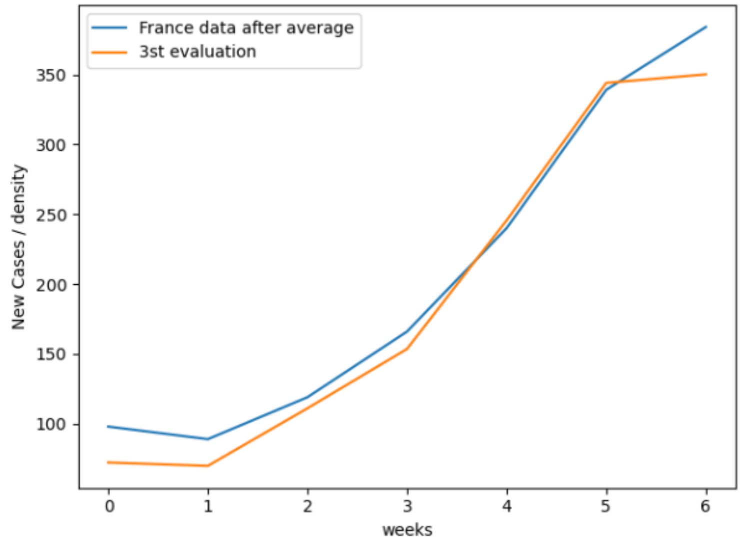


At this point, we noticed that the graph was not as accurate as we hoped it would be.

By looking at the countries data we saw that Belgium had the highest population density compared to France and decided to remove it. Then, we repeated the process once again.

We received that $x = \begin{bmatrix} -2.43433739 \\ -0.43562193 \\ 1.45005519 \\ 0.46228411 \\ 9.2782457 \\ 1.06912461 \end{bmatrix}$ with a minimal error of: 333.052

We received the following graph which is quite close to the original data:



At the end, we calculated the ratio between our model error and the real data and received that the ratio is: **0.00087826**.