1) a. l_2 norm : $\min_c \{(c-x_1)^2 + (c-x_2)^2 + (c-x_3)^2\}$

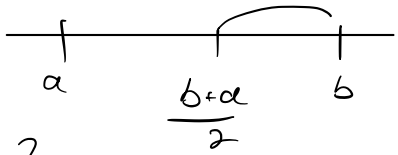
$$\frac{df}{dc} = 2(c-x_1) + 2(c-x_2) + 2(c-x_3) = 6c - 2x_1 - 2x_2 - 2x_3$$

$$6c = 2x_1 + 2x_2 + 2x_3$$

$$c = \frac{x_1 + x_2 + x_3}{3}$$

b. l_∞ norm: $\min_c \{\max(|c-x_1|, |c-x_2|, |c-x_3|)$

chose: $c = \frac{x_3+x_1}{2}$



$$\max\left\{\left|\frac{x_3+x_1}{2} - x_1\right|, \left|\frac{x_3+x_1}{2} - x_2\right|, \left|\frac{x_3+x_1}{2} - x_3\right|\right\}$$

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

נראה כשלכל $c' \neq \frac{x_3+x_1}{2}$ מתקיים $\max\{\ \} >$ -

מקרה 1: $c' \notin [x_1, x_3]$, נניח כי $c' < x_1$:

$$\max\{|c'-x_1|, |c'-x_2|, |c'-x_3|\} = |c'-x_3| > |x_1 - x_3| > \frac{x_1+x_3}{2}$$

אפשרות 2: $x_1 \le c' < \frac{x_3+x_1}{2}$

$$\max\{|c'-x_1|, |c'-x_2|, |c'-x_3|\} = |c'-x_3| > \frac{x_1+x_3}{2}$$

אפשרות 3: $\frac{x_3+x_1}{2} < c' \le x_3$

$$\max\{|c'-x_1|, |c'-x_2|, |c'-x_3|\} = |c'-x_1| > \frac{x_1+x_3}{2}$$

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

chose: $c = x_2$ $\qquad x_1 < x_2 < x_3$

than ℓ_1 norm $= |x_2 - x_1| + |x_2 - x_2| + |x_2 - x_3| = x_2 - x_1 - (x_2 - x_3)$

$= x_3 - x_1$

נניח בשלילה שקיימת נק' c' כך שהסכום יותר קטן

ניתן לחלק למקרים:

$c' \in [x_1, x_2)$: אז -

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

$= c' - x_1 - c' + x_2 - c' + x_3 = x_3 - x_1 + (x_2 - c') > x_3 - x_1$

$c' \in [x_2, x_3)$:

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

$= x_3 - x_1 + (c' - x_2) \geq x_3 - x_1$

$c' \notin (x_1, x_3)$:

• $c' \leq x_1$:

$|c' - x_1| + |c' - x_2| + |c' - x_3| \geq x_3 - x_1$

• $c' \geq x_3$

$|c' - x_1| + |c' - x_2| + |c' - x_3| \geq x_3 - x_1$

② a. סימטריות: $(A^TA)^T = A^T \cdot (A^T)^T = A^TA$ ✓

חיוביות: $x^TA^TAx = (Ax)^T \cdot (Ax) = \langle Ax, Ax \rangle \geq 0$

מ.פ ≥ 0

b. $Cv = \lambda v \Rightarrow (I-C)\cdot v = Iv - Cv = v - \lambda v = (1-\lambda)v$

$\Downarrow$

$(1-\lambda)$ ע"ע של $I-C$ עם וקטור עצמי v.

c. $\operatorname{rank}(A) = n \iff \operatorname{rank}(A^TA) = n$

$\operatorname{rank}(A)=n$ תהי

יהי: $x \in \ker(A)$, מתקיים $Ax = 0$

$x \in \ker(A^TA)$: נכיח

$A^TAx = A^T \cdot (Ax) = A^T \cdot 0 = 0$

יהי: $x \in \ker(A^TA)$, מתקיים $A^TAx = 0$ (נכפיל ב-x^T משמאל)

$x^TA^TAx = 0 \Rightarrow \langle Ax, Ax \rangle = 0 \Rightarrow Ax = 0$

לכן $\ker(A^TA) = \ker(A)$

הסיכום $A^TA \Leftarrow \dim(\operatorname{rank}(A^TA)) = n \Leftarrow n = \dim(A^TA)$

$\Rightarrow$: נשח A^TA הפיכה, לכן $\text{rank}(A^TA)=n \Rightarrow \dim\ker(A^TA)=0$

יהי $x\in\ker(A) \Rightarrow Ax=0$ ← נכפול ב-A^T משמאל $\Rightarrow A^TAx=0 \Rightarrow$

$x\in\ker(A^TA) \Rightarrow \ker(A)\subseteq\ker(A^TA)$

אבל $\ker(A^TA)=\{0\}$ לכן $\ker(A)=\{0\} \Rightarrow \text{rank}(A)=n$

∎

(d) Suppose that $A\in\mathbb{R}^{m\times n}$, $m\geq n$. Show that A is full rank if and only if $A^\top A$ is symmetric and positive definite (you can use the previous section).

$\text{rank}(A)=n \iff A^TA$ חיובית ממש

אנחנו יודעים מסעיפים קודמים ש- A^TA חיובית, סימטרית והפיכה.

כלומר לכל x $A^TAx\geq 0$ אבל $\ker(A^TA)=\{0\}$ לכן $A^TAx>0$

מצד שני, אם A^TA חיובית ממש ד״כ הפיכה ולפי סעיף (c) אז A מדרגה מלאה.

ט

צ"ל $<(A^TA+\alpha I)v, v> \,>0$ לכל $0\neq v\in V$

ידוע כי המטריצה $A^TA\geq 0$ מוחלט לכל $<A^TAv, v> \,\geq 0$

$$<(A^TA+\alpha I)v, v> \;=\; <A^TAv+\alpha Iv, v> \;=\; <A^TAv, v> + <\alpha Iv, v>$$

$$= \underset{\substack{\vee\!\!|\\0}}{<A^TAv, v>} + \alpha \underset{\substack{\vee\\0}}{<v, v>} \Rightarrow \;>0$$

נראה ש- $A^TA+\alpha I$ סימטרית:

בסעיף (α) הראנו ש A^TA סימטרית, αI סימטרית גם כן.

סכום של מטריצות סימטריות זו מטריצה סימטרית

לכן $A^TA+\alpha I$ סימטרית.

(a) Find the best approximation in a least square sense for the system $A\mathbf{x} \approx \mathbf{b}$ where

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \quad , \mathbf{b} = \begin{bmatrix} 6 \\ 1 \\ 5 \\ 2 \end{bmatrix}. \tag{1}$$

Write the normal equations and solve the problem using a computer. You may use built-in functions and provide the code.

$$find(x) \rightarrow \min \|Ax - b\|_2^2 \quad , f(x) = \|Ax-b\|_2^2 = (Ax-b)^T(Ax-b)$$

$$= x^T A^T Ax - 2b^T Ax + b^T b$$

$$\nabla f(x) = 2A^T Ax - 2A^T b = 0$$

$$2A^T Ax = 2A^T b \quad |:2$$

$$A^T Ax = A^T b \qquad \text{Normal Equation}$$

$$x^* = \begin{bmatrix} 1.7 \\ 0.6 \\ 0.7 \end{bmatrix}$$

:הפתרון הנכון הוא

הקוד:

```
a = np.array([[2,1,2],[1,-2,1],[1,2,3],[1,1,1]])
b = np.array([6,1,5,2])
x = numpy.linalg.lstsq(a,b)[0]
```

[2]
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
a=np.array([[2,1,2],[1,-2,1],[1,2,3],[1,1,1]])
b =np.array([6,1,5,2])
x = numpy.linalg.lstsq(a,b)

print(x)
```

```
(array([1.7, 0.6, 0.7]), array([1.4]), 3, array([4.96978399, 2.53339065, 0.93977598]))
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: `rcond` parameter will change to the default of machine precision times ``max(M, N)
To use the future default and silence this warning we advise to pass `rcond=None`, to keep using the old, explicitly pass `rcond=-1`.
  This is separate from the ipykernel package so we can avoid doing imports until
```

(b) Is the solution $\mathbf{x}^*$ that you found in the previous section unique? Explain. What is the minimal objective (loss) value $\|A\mathbf{x}^* - \mathbf{b}\|_2^2$?

A מדרגה מלאה (3) לכן A^TA הפיכה (סעיף c)
לכן ניתן לפתור באופן הבא:

$$A^TAx = A^Tb \quad \Big| \text{ נכפיל ב- } (A^TA)^{-1} \text{ משמאל}$$

$$(A^TA)^{-1}(A^TA)x = (A^TA)^{-1}A^Tb$$

$$x^* = (A^TA)^{-1}A^Tb \longleftarrow$$ זה ה-x עבורו הנורמה מינימלית, כלומר x "יחודי".

<u>נחשב את השגיאה המינימלית:</u>

$$\|Ax^* - b\|_2^2 = \left\| \begin{pmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1.7 \\ 0.6 \\ 0.7 \end{pmatrix} - \begin{pmatrix} 6 \\ 1 \\ 5 \\ 2 \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} \frac{27}{5} \\ \frac{6}{5} \\ 5 \\ 3 \end{pmatrix} - \begin{pmatrix} 6 \\ 1 \\ 5 \\ 2 \end{pmatrix} \right\|_2^2$$

$$= \left\| \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 0 \\ 1 \end{pmatrix} \right\|_2^2 = \frac{7}{5}$$

(c) Find the least squares solution of the system in Eq. (1), but now find a solution for which the second equation is almost exactly satisfied (let's say, such that $|r_2| < 10^{-3}$). Hint: use weighted least squares.

נפתור את המערכת באמצעות Weighted least Squares

נמצא $\hat{x} = (A^T W A)^{-1} A^T W b$ כך ש- $|r_2| < 10^{-3}$

נבחר $W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$$\hat{x} = \left(\begin{pmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \right)^{-1} \cdot \begin{pmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 1 \\ 5 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1160003}{68002} \\ \frac{229994}{340001} \\ \frac{440005}{68002} \end{pmatrix} \Rightarrow r = A\hat{x} - b = \begin{pmatrix} \frac{-210000}{340001} \\ \frac{7}{340001} \leftarrow r_2 \\ 0 \\ \frac{350000}{340001} \end{pmatrix}$$

$$\frac{7}{340001} = r_2 < 10^{-3}$$

(d) Find the least squares solution of the system in Eq. (1), but now add simple Tikhonov regularization term $\lambda\|\mathbf{x}\|_2^2$ with $\lambda = 0.5$.

: Tikhonov regularization נוסחת

$$\hat{x} = \underset{x \in \mathbb{R}^n}{\arg\min} \left\| \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

$$\Downarrow$$

$$\hat{x} = (A^T A + \lambda I)^{-1} \cdot A^T b$$

$$A = \begin{pmatrix} 2 & 1 & 2 \\ 1 & -2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}, b = \begin{pmatrix} 6 \\ 1 \\ 5 \\ 2 \end{pmatrix}, \lambda = 0.5$$ נתון $\hat{x}$ נוסחת

$$\hat{x} = \begin{pmatrix} \frac{2672}{1929} \\ \frac{344}{643} \\ \frac{572}{643} \end{pmatrix}$$

4. **Frobenius Norm.**
See the definition of Frobenius Norm:
`https://mathworld.wolfram.com/FrobeniusNorm.html`
This norm is often used to compare data sets, or apply manipulations on data sets. It is easy to see that by definition

$$\|A\|_F^2 = \sum_{i,j}(a_{ij})^2 = \sum_j \|\mathbf{a}_j\|_2^2$$

where $\mathbf{a}_j$ is a column of A.

(a) Suppose that we want to solve the problem

$$\arg\min_{X\in\mathbb{R}^{n\times n}} \|AX - B\|_F^2$$

where $A, B, X \in \mathbb{R}^{n\times n}$. Find an expression for the solution. When is the solution unique? Hint: this is almost the same as standard least squares.

$$f(X) = \|AX - B\|_F^2 = \sum_j \|Ax_j - b_j\|_2^2 = \sum_j (Ax_j - b_j)^T(Ax_j - b_j)$$

x_j העמודה ה-j של X.

$$= \sum_j (x_j^T A^T - b_j^T)(Ax_j - b_j) = \sum_j x_j^T A^T A x_j - x_j^T A^T b_j - b_j^T A x_j + b_j^T b_j$$

$$\nabla f(X) = \begin{pmatrix} 2A^TAx_1 - 2A^Tb_1 \\ 2A^TAx_2 - 2A^Tb_2 \\ \vdots \\ 2A^TAx_n - 2A^Tb_n \end{pmatrix} = 0$$

לכל j (מ-1 עד n)

$$2A^TAx_j - 2A^Tb_j = 0$$

נניח ש-A הפיכה:

$$A^TAx_j = A^Tb_j \quad \text{(הכפלה ב-} (A^T)^{-1} \text{ משמאל)}$$

$$x_j = (A^T)^{-1}A^Tb_j \quad \text{(הכפלה ב-} A^{-1} \text{ משמאל)}$$

$x_j = A^{-1} b_j$

מעיקרון כל x_j הוא "חוף", כלומר x_j "חוף" על A הפסכה .

(b) Often, we wish to make data A look similar to data B of the same size, to remove artifacts that are not related to the phenomena that we test. For example, when we measure gene expression values, often there's a bias of the particular lab's measurement. To correct that, we to make the datasets similar, and one option is to solve a minimization of the following form (this is a bit simplified)

$$\underset{D\in\mathbb{R}^{n\times n}, D \text{ is diagonal}}{\arg\min} \|DA - B\|_F^2$$

D is a diagonal matrix. In other words, we wish to find a scale D_{ii} for each row of A (denoted by $\mathbf{a}_i$) so that $\|\mathbf{a}_i D_{ii} - \mathbf{b}_i\|_2^2$ is minimized ($\mathbf{b}_i$ is a row in B). Use a computer and find the solution D for the following matrices:

$$A = \begin{bmatrix} 5 & 6 & 7 & 8 \\ 1 & 3 & 5 & 4 \\ 1 & 0.5 & 4 & 2 \\ 3 & 4 & 3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 0.57 & 0.56 & 0.8 & 1 \\ 1.5 & 4 & 6.7 & 4.9 \\ 0.2 & 0.1 & 1 & 0.6 \\ 11 & 30 & 26 & 10 \end{bmatrix}$$

$$x_i = \frac{a_i b_i^T}{a_i a_i^T}$$

$$D = \begin{pmatrix} x_1 & & & 0 \\ & x_2 & & \\ & & x_3 & \\ 0 & & & x_4 \end{pmatrix} \qquad D_{ii} = x_i \quad \text{לכן}$$

$$f(x_i) = \|a_i x_i - b_i\|_2^2 = (a_i x_i - b_i)(a_i x_i - b_i)^T = (a_i x_i - b_i)(x_i^T a_i^T - b_i^T)$$

$$= a_i x_i x_i^T a_i^T - a_i x_i b_i^T - b_i x_i^T a_i^T + b_i b_i^T$$

$$f'(x_i) = 2x_i a_i a_i^T - 2a_i b_i^T = 0$$

$$x_i a_i a_i^T = a_i b_i^T$$

$$\boxed{x_i = \frac{a_i b_i^T}{a_i a_i^T}}$$

$$x_1 = \frac{\frac{1981}{100}}{174} = 0.113$$

$$x_2 = \frac{\frac{333}{5}}{51} = 1.305$$

$$x_4 = \frac{241}{35} = 6.885$$

$$x_3 = \frac{\frac{109}{20}}{21.25} = 0.256$$

$$D = \begin{pmatrix} 0.113 & & & \\ & 1.305 & & 0 \\ & & 0.256 & \\ & 0 & & 6.885 \end{pmatrix}$$

سؤال 5:

Code:

```
import matplotlib
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression

path = 'insurData.csv'
df = pd.read_csv(path)
print('\nNumber of rows and columns in the data set: ', df.shape)
print('')
df.head()

# a
df['sex'] = df['sex'].replace(['female'], 0)
df['sex'] = df['sex'].replace(['male'], 1)
df['smoker'] = df['smoker'].replace(['yes'], 1)
df['smoker'] = df['smoker'].replace(['no'], 0)
df['region'] = df['region'].replace(['northeast'], 0)
df['region'] = df['region'].replace(['southeast'], 1)
df['region'] = df['region'].replace(['southwest'], 2)
df['region'] = df['region'].replace(['northwest'], 3)

# b
df['charges'] = df['charges'] / 1000
meanc = df['charges'].mean()
df['charges'] = df['charges'] - meanc
meanb = df['bmi'].mean()
df['bmi'] = df['bmi'] - meanb
meana = df['age'].mean()
df['age'] = df['age'] - meana
```

```
# c
for i in range(5):
    train, test = train_test_split(df, test_size=0.2)
    train_x = train[['age', 'sex', 'bmi', 'children', 'smoker',
'region']].to_numpy()
    train_y = train[['charges']].to_numpy()
    test_x = test[['age', 'sex', 'bmi', 'children', 'smoker',
'region']].to_numpy()
    test_y = test[['charges']].to_numpy()

    ans_train = np.linalg.lstsq(train_x, train_y, rcond=None)[0]
    train_mse = (1 / np.size(train_x, 0)) * (
            (np.linalg.norm(train_x @ ans_train - train_y, ord=2,
axis=None, keepdims=False)) * (np.linalg.norm(train_x @ ans_train -
train_y, ord=2, axis=None, keepdims=False)))
    print('train MSE is: ')
    print(train_mse)
    ans_test = np.linalg.lstsq(test_x, test_y, rcond=None)[0]
    test_mse = (1 / np.size(test_x, 0)) * (
            (np.linalg.norm(test_x @ ans_test -
test_y,ord=2,axis=None,keepdims=False)) * (np.linalg.norm(test_x @
ans_test - test_y, ord=2, axis=None, keepdims=False)))
    print('test MSE is: ')
    print(test_mse)
    compered_mse = train_mse - test_mse
    print('diff is: ')
    print(compered_mse)

#d
matplotlib.pyplot.hist(train_x @ ans_train - train_y,bins=100)
plt.show()
```

<u>התייחסות לסעיף ג'</u>: קיבלנו שהשגיאה הממוצעת היא באיזור ה40+ (כלומר בפועל פי 1000), זו תוצאה לא טובה כלומר המודל לא צופה טוב את החיובים.

Results :

Number of rows and columns in the data set: (1338, 7)

///b

	age	sex	bmi	children	smoker	region	charges
0	-20.207025	0	-2.763397	0	1	2	3.614502
1	-21.207025	1	3.106603	1	0	1	-11.544870
2	-11.207025	1	2.336603	3	0	1	-8.820960
3	-6.207025	1	-7.958397	0	0	3	8.714048
4	-7.207025	1	-1.783397	0	0	3	-9.403567
...	...	...	...	...	...	...	...
1333	10.792975	1	0.306603	3	0	3	-2.669874
1334	-21.207025	0	1.256603	0	0	0	-11.064441
1335	-21.207025	0	6.186603	0	0	1	-11.640589
1336	-18.207025	0	-4.863397	0	0	2	-11.262477
1337	21.792975	0	-1.593397	0	1	3	15.870938

///c

train MSE is:

40.782937519716704

test MSE is:

45.84063733557119

diff is:

-5.057699815854484

train MSE is:

41.73079311544873

test MSE is:

42.877532999891436

diff is:

-1.1467398844427024

train MSE is:

41.33098523038746

test MSE is:

44.93500684381962

diff is:

-3.6040216134321597

train MSE is:

40.44832886730718

test MSE is:

47.83190326975075

diff is:

-7.383574402443564

train MSE is:

43.30836210121807

test MSE is:

37.15719576892592

diff is:

6.151166332292156

////d