

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN



TOÁN ỨNG DỤNG THỐNG KÊ

BÁO CÁO THỰC HÀNH

LAB 03

Mã số sinh viên: 21120582

Họ và Tên: Đinh Hoàng Trung.

Mail: 21120582@student.hcmus.edu.vn.

MSSV: 21120582

Họ Tên: Đinh Hoàng Trung.

1. Khái quát bài làm:

- Bài tập được làm trên Visual Studio Code.
- Ma trận đầu vào sẽ được nhập vào file .txt.
 - o Ví dụ:

```
data.txt
1 1971 2250
2 1972 2500
3 1974 5000
4 1978 29000
5 1982 120000
6 1985 275000
7 1989 1180000
8 1993 3100000
9 1997 7500000
10 1999 24000000
11 2000 42000000
12 2002 220000000
13 2003 410000000
```

Dữ liệu đọc được:

```
[[1.971e+03 2.250e+03]
 [1.972e+03 2.500e+03]
 [1.974e+03 5.000e+03]
 [1.978e+03 2.900e+04]
 [1.982e+03 1.200e+05]
 [1.985e+03 2.750e+06]
 [1.989e+03 1.180e+07]
 [1.993e+03 3.100e+07]
 [1.997e+03 7.500e+07]
 [1.999e+03 2.400e+08]
 [2.000e+03 4.200e+08]
 [2.002e+03 2.200e+09]
 [2.003e+03 4.100e+09]]
```

- Dữ liệu sau khi được giải nghiệm sẽ xuất ra màn hình terminal (VSC) / console.
- Thư viện được sử dụng : numpy, scipy.linalg, math, matplotlib.pyplot.
- Các hàm theo yêu cầu bài tập:
 - o `least_square_solution(A, B)`: tìm nghiệm bình phương tối thiểu của phương trình $Ax = B$, để tìm hệ số của bài toán
- Các hàm hỗ trợ:
 - o `Readfile_txt()`: đọc dữ liệu từ file data.txt lưu vào ma trận lưu dưới dạng numpy.array.
 - o `log10_model(a)`: chỉnh sửa số liệu phù hợp với mô hình đề bài ($\text{Log}_{10}(N) \approx \theta_1 + \theta_2(t - 1970)$), đặt $x = t - 1970$ và $y = \text{Log}_{10}(N)$. Trả về 2 ma trận A và B theo hệ được nêu trên ý tưởng làm bài.
 - o `modelling_vector(vt)`: mô hình hóa vector vt truyền vào.

MSSV: 21120582

Họ Tên: Đinh Hoàng Trung.

- `find_A_plus(A)`: trả về ma trận giả nghịch đảo của ma trận `A` truyền vào.
- `inh_y(x,a,b)`: trả về giá trị $y = a*x + b$.

2. Giải thích thuật toán và ý tưởng làm bài:

- Mô hình theo đề bài:

$$\log_{10} N \approx \theta_2(t - 1970) + \theta_1$$

- Theo đó dữ liệu được cung cấp là t và N , trong đó t là biến độc lập thể hiện năm ra đời của một bộ vi xử lí, N là biến phụ thuộc vào t thể hiện số bóng bán dẫn trong bộ vi xử lí ra đời vào năm t .
- Đặt $y = \log_{10}(N)$ và $x = (t - 1970)$, từ đó mô hình sẽ chuyển về dạng $y \approx \theta_2 x + \theta_1$.

- Với N và t từ số liệu ta lập hệ phương trình:

$$\text{Log}_{10}(2250) = a * (1971 - 1970) + b$$

$$\text{Log}_{10}(2500) = a * (1972 - 1970) + b$$

$$\text{Log}_{10}(5000) = a * (1974 - 1970) + b$$

$$\text{Log}_{10}(29000) = a * (1978 - 1970) + b$$

$$\text{Log}_{10}(120000) = a * (1982 - 1970) + b$$

$$\text{Log}_{10}(275000) = a * (1985 - 1970) + b$$

.....

$$\text{Log}_{10}(410000000) = a * (2003 - 1970) + b$$

$$\Rightarrow \text{Ấn cần tìm của hệ là } X = (a, b)$$

Hệ phương trình được biểu diễn dưới dạng $Ax = B$

Với A là ma hệ số của a và b tức các vector $[(t - 1970), 1]$, và B là ma trận hệ số tự do tức các vector $[\text{Log}(N)]$.

Được biểu diễn dưới dạng ma trận như sau:

$$A = [[1, 1],$$

$$[2, 1],$$

$$[4, 1],$$

$$[8, 1],$$

$$[12, 1],$$

$$[15, 1],$$

$$[19, 1],$$

$$[23, 1],$$

$$[27, 1],$$

$$[29, 1],$$

$$[30, 1],$$

MSSV: 21120582

Họ Tên: Đinh Hoàng Trung.

[32, 1],

[33, 1]]

$B = [[3.3521825181113627],$

[3.3979400086720375],

[3.6989700043360187],

[4.462397997898956],

[5.079181246047625],

[5.4393326938302625],

[6.0718820073061255],

[6.4913616938342725],

[6.8750612633917],

[7.380211241711606],

[7.6232492903979],

[8.342422680822206],

[8.612783856719735]]

$x = [[a],$

[b]]

Theo yêu cầu bài toán thì ta cần phải tìm ra x là nghiệm bình phương tối thiểu của phương trình $Ax = B$:

Theo định lý 1 được cung cấp từ tài liệu: $x = A^+ B$

\Rightarrow Vậy ta cần phải tìm A^+ : ma trận giả nghịch đảo của A .

Ta có $A = U\Sigma V^T$: là phân tích điểm kì dị của ma trận A có kích thước 13×2 được phân tích bằng phân rã SVD.

Ma trận A^+ sẽ được tính như sau: $A^+ = V\Sigma^+U^T$

Từ đó tìm ra $x = A^+ B$, thay vào $\log_{10} N \approx \theta_2(t - 1970) + \theta_1$ được phương trình đường thẳng khớp với mô hình.

3. Câu a:

Lập ma trận từ dữ liệu đọc từ file:

MSSV: 21120582

Họ Tên: Đinh Hoàng Trung.

 data.txt

```
1 1971 2250
2 1972 2500
3 1974 5000
4 1978 29000
5 1982 120000
6 1985 275000
7 1989 1180000
8 1993 3100000
9 1997 7500000
10 1999 24000000
11 2000 42000000
12 2002 220000000
13 2003 410000000
```

Dữ liệu đọc được:

```
[[1.971e+03 2.250e+03]
 [1.972e+03 2.500e+03]
 [1.974e+03 5.000e+03]
 [1.978e+03 2.900e+04]
 [1.982e+03 1.200e+05]
 [1.997e+03 7.500e+06]
 [1.999e+03 2.400e+07]
 [2.000e+03 4.200e+07]
 [2.002e+03 2.200e+08]
 [2.003e+03 4.100e+08]]
```

Ma Trận A, B trong phương trình $Ax = B$:

```
- A = [[1, 1],
       [2, 1],
       [4, 1],
       [8, 1],
       [12, 1],
       [15, 1],
       [19, 1],
       [23, 1],
       [27, 1],
       [29, 1],
       [30, 1],
       [32, 1],
       [33, 1]]
- B = [[3.3521825181113627],
       [3.3979400086720375],
       [3.6989700043360187],
       [4.462397997898956],
       [5.079181246047625],
       [5.4393326938302625],
       [6.0718820073061255],
       [6.4913616938342725],
       [6.8750612633917],
       [7.380211241711606],
       [7.6232492903979],
       [8.342422680822206],
```

MSSV: 21120582

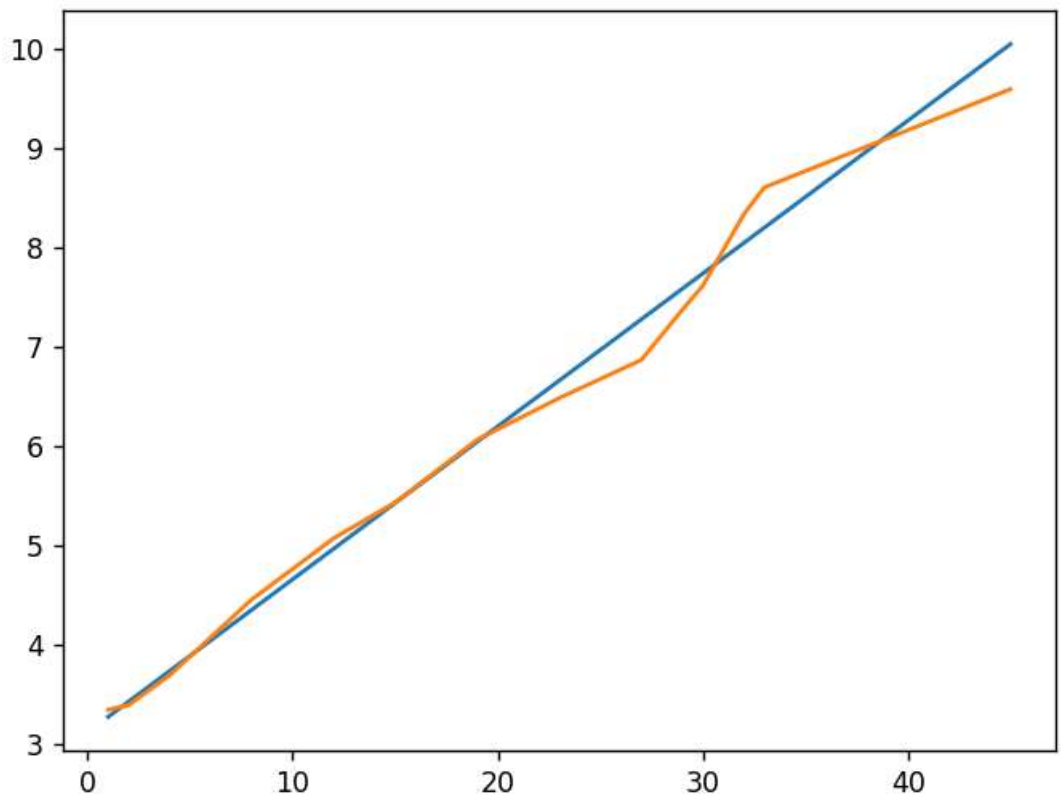
Họ Tên: Đinh Hoàng Trung.

[8.612783856719735]]

Tìm nghiệm bình phương tối tiểu:

- Tìm A^+ : bằng công thức $A^+ = V\Sigma^+U^T$
 - Phân rã SVD ma trận A nhận được U , các giá trị riêng $\sqrt{\sigma}$, và ma trận V .
 - Chuyển đổi các giá trị riêng $\sqrt{\sigma}$ thành ma trận Σ^+ .
 - Tìm $A^+ = V\Sigma^+U^T$.
 - $x = A^+ B$.
- Có $x = (a, b)$: viết phương trình đường thẳng khớp với mô hình.

$$y = [0.15401818] x + [3.12559263]$$



4. Câu b:

Từ câu a ta có phương trình đường thẳng: $y = [0.15401818] x + [3.12559263]$.

Thay $x = 2015 - 1970$ tìm được $y = \text{Log}_{10}(N)$.

$$N = 10^y = 1.13870369 \cdot 10^{10}$$

MSSV: 21120582

Họ Tên: Đinh Hoàng Trung.

⇒ Số bóng bán dẫn dự đoán trong bộ xử lí giới thiệu vào năm 2015 = $1.13870369.10^{10}$ bóng bán dẫn.

⇒ Chênh lệch với thực tế $7.38703687.10^9$.

----Hết---