EAS 596 Midterm Exam.

Abhishek Kumar #, UB.NO → 50291056

Prof. David Salac.

1. The elliptical path of the planet in a Cartesian Coordinate system is :

$$ay^2 + bxy + cx + dy + e = x^2$$

a) To solve for coefficients a, b, c, d & e, we represent the system as a matrix, for data points (x_i, y_i) as:

$$\underbrace{\begin{bmatrix} y_1^2 & x_1y_1 & x_1 & y_1 & 1 \\ y_2^2 & x_2y_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_m^2 & x_my_m & x_m & y_m & 1 \end{bmatrix}}_{\underline{A}} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_m^2 \end{bmatrix}}_{\underline{b}}$$

where (x_i, y_i) are our data points. It represents $Ax = b$ which can be solved using normal equation or using matlab's "mldivide()" command.

b) Please check "Q1_script.m". (Part b)

c) Please check "Q1_script.m". (Part c)

② a)

i) $A = \begin{bmatrix} 1/3 & 1/3 & 2/3 \\ 2/3 & 2/3 & 4/3 \\ 1/3 & 2/3 & 3/3 \\ 2/5 & 2/5 & 4/5 \\ 3/5 & 1/5 & 4/5 \end{bmatrix}$

reducing A to its rref:

$$\begin{bmatrix} 1/3 & 1/3 & 2/3 \\ 2/3 & 2/3 & 4/3 \\ 1/3 & 2/3 & 3/3 \\ 2/5 & 2/5 & 4/5 \\ 3/5 & 1/5 & 4/5 \end{bmatrix} \xrightarrow[R_3 = R_3 - R_1,\ R_4 = R_4 - \frac{6}{5}R_1]{R_2 = R_2 - 2R_1} \begin{bmatrix} 1/3 & 1/3 & 2/3 \\ 0 & 0 & 0 \\ 0 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ 3/5 & 1/5 & 4/5 \end{bmatrix}$$

$$\Big\downarrow \begin{matrix} R_1 = R_1 \times 3 \\ R_3 = R_3 \times 3 \\ R_5 = R_5 \times 3 \end{matrix}$$

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xleftarrow{R_5 = R_5 + 2R_3} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & -2 & -2 \end{bmatrix} \xleftarrow{R_5 = R_5 - 3R_1} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 3 & 1 & 4 \end{bmatrix}$$

↑ free variable

We see that the last column is a free variable, so the rank$(A) = 2$.

$\therefore$ Exact rank(A) is 2.

iii) To relate this theorem to determination of rank in finite precision mathematics, we can view ε as the machine precision/finite precision and infer from the theorem:

A_ε can be approximated as a low rank matrix A, where where rank$(A) <$ rank(A_ε) where $||A - A_\varepsilon|| < \varepsilon$. That means, if we decrease the finite precision (increase ε), we can get even smaller of rank approximation for A_ε and vice-versa.

2\. b) (i) We see that even by adding minute noise in range [-0.005, 0.005], the orbit has visible change while the points changed slightly. This implies that our coefficient estimation is not stable. This is due to the fact that we have used very less to to predict the coefficients. Also the points were from the same region of the expected ellipse. This made our system unstable.

(ii) We see that for $k=1$, we removed 2 eigenvalues and the resultant orbit was offset significantly from its original orbit. While for $k=2$, we removed 1 eigenvalue and the elliptical orbit changed significantly but less than that with $k=1$.

We see that as we increase the tolerance, the resulting orbit does not fit the data very well.

I would like to prefer a solution that is less sensitive to small perturbations in the data. This is a stable system. The system that fits the data more closely even with less data tends to overfit the data and don't do well in predicting about new data points and suffer from high variance.

3. a) Please check "Q3_senate.m" for part a to e.

b) From the singular values of the senate votes, we see that there are two singular values that have very high magnitude than the rest of the singular values. This indicates of two principle patterns in voting.

c) Please check "Q3_senate.m"

d) We are decomposing matrix A into U, S & V.

~~If~~ Here, the rows of A are people who votes and columns are, let say elections. Then the U matrix ~~represent~~ connects people to concepts and V matrix connects elections to concepts.

Here, lets assume the concepts are topics of the election. Therefore u_1 should represent the

Scanned by CamScanner

the voting pattern in the most dominant category of topic and μ_2 represents the voting pattern of the representatives in the 2nd most dominant category of topics.

(e) We observe that for representatives who don't ~~identify~~ have any preference on the most dominant category of topic ($\mu_1 \sim 0$), we find it difficult to correctly predict their voting.

While for representatives who have strong preference (either very low μ_1 or very high μ_1), we are able to predict their voting pattern with high accuracy.

(9) When we look at the scatter plot for u_1 vs u_2,

for senate data

we observe that for the most dominant topic, ~~i.e.~~

the Republican strongly support it while the Democrats strongly oppose it.

While for the 2nd most domintant topic (u_2), we observe that the Republicans strongly oppose it while the Democrats only mildly oppose it.

We can infer that the Democrats and Republicans have certain views on topics and tend to vote in groups and we can quite accurately predict about their vote on certain topics for which they have strong views (preference i.e high or low U's).

We can infer similarly from house data as well.

For the most dominant category of topic, from the u_1 vs u_2 plot, we see that the Democrats and Republicans clearly oppose each other while for 2nd most dominant category of topic, the Republicans support & mildly-to-strongly while the Democrats support strongly.