2DBL100 Linear Algebra

Dayeong Kim, 1258893
Tae Gun Lee, 1286684
Yeochan Yoon, 1227750
Yi Hang, 1014298

1. Vector angles and movie ratings

a) 8 movie types

8 Sample Set

- Nature movie
- Comedy series
- Romantic comedy
- Science Fiction
- Latest James Bond (action)
- Detective
- Candid camera (Just for laughs / Bananensplit / Verstehen Sie Spass)
- Documentary of historic events

Dayeong - $\{1, 2, 1, 1, -1, 1, 1, -1\}$
Tae Gun -$\{1, 2, 0, 2, 1, 2, 1, 2\}$
Yi Hang - $\{1, 2, 0, 2, 1, 0, -1, 2\}$
Yeochan - $\{-2, 0, 1, 0, 0, -2, 1, 0\}$

b) If we use the scale from 1 to 5 our cosine formula gives a smaller range of angles which may lead a very rough prediction. For the numerator of equation, A * B, the sum of vectors will only grow positively, as vector only contains the positive constants. The consequence is that the angle which ratings can create is limited as it would not cover negative quadrants on the axis.

c) If all movies are rated to 0, comparing other object ratings to this rating would result nothing. Because the norm difference and angle difference are not affected by the zero points, which does not subtract or add anything to it. It would give out the object rating itself.

d) Who is the closest person to rating $\{1, 1, 1, 1, 1, 1, 1, 1, 1\}$?
T has the smallest distance away from the given rating, $\sqrt[2]{5}$. We subtracted each number in vector by 1, then used 2-norm to calculate the length between two points.
T is also the closest to it in terms of angle, which is 26 degree away from it. We multiplied each vector to given vector then calculated length of each of them. Then using the angle formula, T turns out to be the smallest.

e) Y's opinion can be predicted through calculating the magnitude of the vector.
Norm of the vector "Y" of first 7 movies: 3.464
Norm of the vector "Y" including the unknown 8th movie (denoted as x) varies from:
3.464 (x = 0), 3.605 (x = 1, -1) and 4. (x = 2, -2)
Another method of predicting Y's opinion is comparing the angle of Y with X's opinion who rated all movies to 1 in the previous question.
Y = { 1, -1, 0, 2, -2, 1, 1, x }
X = $\{1, 1, 1, 1, 1, 1, 1, 1\}$
cos(X, Y) = $\frac{2+x}{(\sqrt[2]{x^2+12})(\sqrt[2]{8})}$
The angle varies from 78.221 (x=0)
72.89 (x=1)
69.295 (x=2)
Therefore, Y's opinion is similar to Yi Hang's rating in terms of the magnitude of the vector since the rating of Yi Hang's vector was 3.87. In terms of angle,

none of our member had a relevant angle since the angle of Y presented around 73.468 throughout the mean of three angles measured.

2. Google Page rank

a) In our example we have created the rank of 12 teams in Korean football league in 2017. Each team played 22 matches.

Connectivity matrix G – Korea football league 2017

Home \ Away	DGU	GWN	GWJ	ICU	JJU	JHM	JND	PHS	SJS	SEO	SSB	USH
Daegu FC		1–2	2–2	2–2	1–2	0–2	2–1	3–0	2–0	2–1	0–3	1–3
Gangwon FC	2–1		2–2	2–1	2–1	1–1	2–2	2–2	2–0	0–1	1–2	0–1
Gwangju FC	1–0	1–1		0–0	0–1	1–0	1–2	1–2	1–2	3–2	0–3	1–1
Incheon United	0–0	1–1	1–0		0–1	0–0	1–3	0–3	1–1	1–5	3–3	1–2
Jeju United	4–2	1–2	1–1	1–1		2–1	2–0	3–0	4–1	1–2	1–2	3–0
Jeonbuk Hyundai Motors	2–2	4–1	3–1	1–1	0–4		2–1	2–0	4–1	1–0	2–0	4–0
Jeonnam Dragons	4–3	2–1	5–0	3–2	2–2	0–3		1–3	1–3	2–2	1–3	5–0
Pohang Steelers	2–1	1–2	2–0	2–0	2–1	1–3	1–1		1–2	3–2	2–3	1–2
Sangju Sangmu	1–1	1–2	1–0	0–1	0–3	1–3	1–1	0–1		2–2	1–3	0–0
FC Seoul	0–0	2–3	2–1	3–0	0–0	2–1	1–0	1–0	1–2		1–1	0–0
Suwon Samsung Bluewings	1–1	3–3	0–0	3–0	1–0	0–2	4–1	1–0	0–0	1–2		1–2
Ulsan Hyundai	1–0	2–1	1–0	1–2	1–0	0–0	1–0	2–1	0–1	1–1	2–1	

Home \ Away	DGU	GWN	GWJ	ICU	JJU	JHM	JND	PHS	SJS	SEO	SSB	USH
Daegu FC		0	0	0	0	0	0.1	1/11	0.1	1/6	0	0
Gangwon FC	0.1		0	0.1	2/7	0	0	1/11	0.2	1/6	0	0
Gwangju FC	0.1	0		0	0	0.25	0	0	0	1/6	0	0
Incheon United	0	0	1/12		0	0	0	0	0.1	0	0	0.2
Jeju United	0.2	0	1/12	0.1		0.5	0.1	1/11	0.2	0	0	0.2
Jeonbuk Hyundai Motors	0.1	1/6	1/12	0	0		0.2	2/11	0.2	1/6	0.4	0.2
Jeonnam Dragons	0.1	1/6	2/12	0.2	0	0		0	0	0	0	0.2
Pohang Steelers	0.1	0	2/12	0.2	1/7	0	0.1		0.1	1/6	0	0
Sangju Sangmu	0	0	2/12	0	0	0	0.1	1/11		1/6	0	0.2
FC Seoul	0	1/6	1/12	0.2	1/7	0.25	0.1	1/11	0		0.2	0
Suwon Samsung Bluewings	0.1	1/6	1/12	0.1	2/7	0	0.2	2/11	0.1	0		0
Ulsan Hyundai	0.2	2/6	1/12	0.1	1/7	0	0.1	2/11	0	0	0.4	

Home \ Away	DGU	GWN	GWJ	ICU	JJU	JHM	JND	PHS	SJS	SEO	SSB	USH
Daegu FC		0 0.0125	0 0.0125	0 0.0125	0 0.0125	0 0.0125	0.1 0.0975	1/11 0.0898	0.1 0.0975	1/6 0.1542	0 0.0125	0 0.0125
Gangwon FC	0.1 0.0975		0 0.0125	0.1 0.0975	2/7 0.2554	0 0.0125	0 0.0125	1/11 0.0898	0.2 0.1825	1/6 0.1542	0 0.0125	0 0.0125
Gwangju FC	0.1 0.0975	0 0.0125		0 0.0125	0 0.0125	0.25 0.225	0 0.0125	0 0.0125	0 0.0125	1/6 0.1542	0 0.0125	0 0.0125
Incheon United	0 0.0125	0 0.0125	1/12 0.0833		0 0.0125	0 0.0125	0 0.0125	0 0.0125	0.1 0.0975	0 0.0125	0 0.0125	0.2 0.1825
Jeju United	0.2 0.1825	0 0.0125	1/12 0.0833	0.1 0.0975		0.5 0.4375	0.1 0.0975	1/11 0.0898	0.2 0.1825	0 0.0125	0 0.0125	0.2 0.1825
Jeonbuk Hyundai Motors	0.1 0.0975	1/6 0.1542	1/12 0.0833	0 0.0125	0 0.0125		0.2 0.1825	2/11 0.1670	0.2 0.1825	1/6 0.1542	0.4 0.3525	0.2 0.1825
Jeonnam Dragons	0.1 0.0975	1/6 0.1542	2/12 0.1542	0.2 0.1825	0 0.0125	0 0.0125		0 0.0125	0 0.0125	0 0.0125	0 0.0125	0.2 0.1825
Pohang Steelers	0.1 0.0975	0 0.0125	2/12 0.1542	0.2 0.1825	1/7 0.1339	0 0.0125	0.1 0.0975		0.1 0.0975	1/6 0.1542	0 0.0125	0 0.0125
Sangju Sangmu	0 0.0125	0 0.0125	2/12 0.1542	0 0.0125	0 0.0125	0 0.0125	0.1 0.0975	1/11 0.0898		1/6 0.1542	0 0.0125	0.2 0.1825
FC Seoul	0 0.0125	1/6 0.1542	1/12 0.0833	0.2 0.1825	1/7 0.1339	0.25 0.225	0.1 0.0975	1/11 0.0898	0 0.0125		0.2 0.1825	0 0.0125
Suwon Samsung Bluewings	0.1 0.0975	1/6 0.1542	1/12 0.0833	0.1 0.0975	2/7 0.2554	0 0.0125	0.2 0.1825	2/11 0.1670	0.1 0.0975	0 0.0125		0 0.0125
Ulsan Hyundai	0.2 0.1825	2/6 0.1542	1/12 0.0833	0.1 0.0975	1/7 0.1339	0 0.0125	0.1 0.0975	2/11 0.1670	0 0.0125	0 0.0125	0.4 0.3525	

b) The following is the eigenvector:

0.04

0.08

0.06
0.04
0.13
0.13
0.06
0.07
0.07
0.11
0.10
0.11

c) The above eigenvector corresponds to chance of winning per each team such that..
DGU: 4%
GWN: 8%
GWJ: 6%
ICU: 4%
JJU: 13%
JHM: 13%
JND: 6%
PHS: 7%
SJS: 7%
SEO: 11%
SSB: 10%
USH: 11%
This gives a rank such that..
1. JHM: 13% and JJU: 13%
2.
3. SEO: 11% and USH: 11%
4.
5. SSB: 10%

6. GWN: 8%
7. PHS: 7% and SJS: 7%
8.
9. GWJ: 6% and JND: 6%
10.
11. DGU: 4% and ICU: 4%
12.

The resulted rank contains 5 duplicate rankings, which seem a bit odd. This suggests some calculation flaws in the process of getting the results. The actual ranking after 22 matches was like this.

1. JHM
2. USH
3. SSB
4. JJU
5. GWN
6. SEO
7. PHS
8. JND
9. SJS
10. DGU
11. ICU
12. GWJ

This shows 2 correctly matched ranks for rank 1 and rank 12. Other rankings from our calculation are a bit off the actual result, this suggests our result is not that reliable. Possible reasons for unreliable results include; incorrectly assigned points to data and calculation mistakes.

d) If p=0.99, the pagerank would result a team with many wins to take the top of the league by a larger chance, making other teams with less wins only have a very small chance.

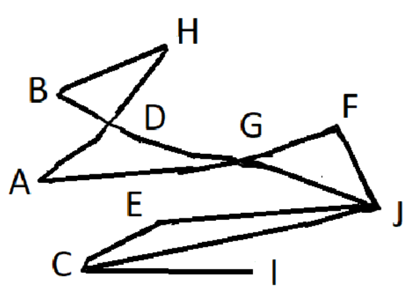
If p=0.5, the pagerank would result each team to have a similar probability that they would have a small difference in probability to be ranked up and down.

e) Our first node will be team DGU, which is ranked at 1-th place. We can maximize its rank by taking a link from a node with the most links. This will affect the rank because it would have more probability from other nodes to link to our node.

3. Clustering

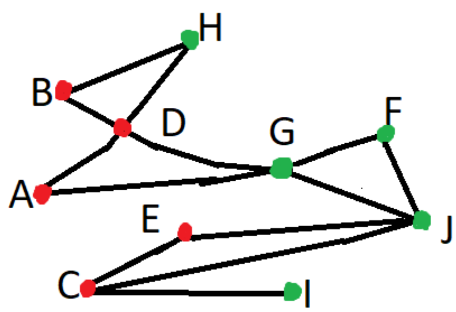
a) Our group chose to make a graph on friends on Facebook. We have 10 nodes as 10 people and nodes are connected with a line if they are friends on Facebook.

b)



d and e)

Eigenvector	Fiedler vector
-0.21696014	-0.41242477
-0.41242477	-0.41242477
0.36409516	-0.29966846
-0.29966846	-0.21696014
0.31385013	-0.07493525
0.05957582	0.05957582
-0.07493525	0.17779896
-0.41242477	0.31385013
0.50109331	0.36409516
0.17779896	0.50109331



Coloring the node given by the Fiedler vector does not give a clear indication of clusters, where node like H which we expected to be red was actually green. There exist some odd points, which was an unexpected result.

4. Data mining: term-document matrix

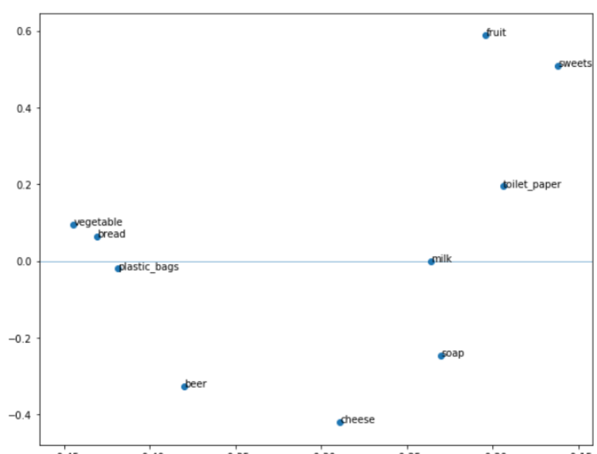
a) We chose to show the shopping list of 10 customers over 10 items. Everyone bought at least one item.

b)

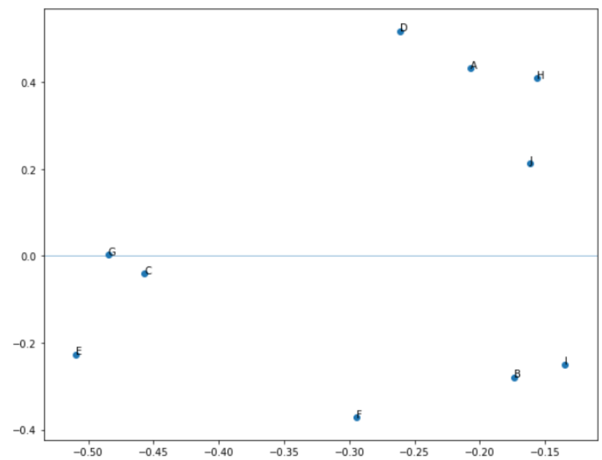
	A	B	C	D	E	F	G	H	I	J	Total
Bread	1		1		1		1				4
Fruit	1			1				1		1	4
Milk			1			1		1			3
Cheese		1			1	1			1		4
Vegetable			1	1	1		1				4
Beer		1			1	1	1				4
Sweets	1			1				1			3
Toilet Paper				1			1				2
Soap			1			1			1		3
Plastic Bags			1		1		1			1	4
Total	3	2	5	4	5	4	5	3	2	2	

d)

Product graph (u1 horizontal and u2 vertical)



Customer graph (v1 horizontal and v2 vertical)



e) We expected that products bought the most times by customers would have high u1 but some products like beer for example, had a low u1. This suggests that important products will affect the shape of graph the most, which does not rely on the number of items bought.
The importance of customers was surprising because since everyone has bought max 4 products, it was tough to determine who's got the many important products. It turned out to be D who is the most important customer. This is because D bought 4 products that are all above 0.0 in u1.

5. Speed and memory of your laptop

a) We used the code: start = time.time(); A+B; end = time.time()
For 5 tries we get...
1) 1736433144524.6406
2) 1572251221573.5498
3) 1646810444955.1277
4) 1595350983753.3215
5) 1636257668643.13
The average gives 1637420692689.9539 flops.

b) We used the code: start = time.time(); dot(A,B); end = time.time()

For 5 tries we get...

1) 126426724568.51909

2) 131237016693.37114

3) 126607774434.19606

4) 124024547711.52573

5) 128916856419.20982

The average gives 127442583965.36438 flops.

c) In general, addition takes about 10 times more flops than multiplications.

d) Everyone tried 5 times and get the average.

Addtion

Da Young: 1840121961074.1733

Taegun: 1263766807301.2954

Yi: 913854724685.7937

Yeochan: 1627458692689.4657

Multiplication

D: 1861382580056.1094

T: 72511487123.87018

Yi: 72982184540.89731

Ye: 125639583537.73640

For addition, Yi's laptop was the fastest to compute. For multiplication, Taegun's laptop was the fastest. The maximum minus minimum flops in addition is 926267236388.3796, for multiplication the difference is 1788871092932.2393, which is twice bigger than difference of addition.

e) when n = 24000, it gives an memory error on A+B.

6. Vector and matrix norms

a)
1. $|x| \geq$ 0
2. $|x| =$ x $\leftrightarrow$ x = 0
3. $|ax| = |a|$*$|x|$ if a $\in$ R
4. $|x + y| \leq |x| + |y|$

b)
1. if $|x|$ contains at least one number != 0, then 1-norm != 0.
2. if 1-norm = 0, then all elements $\in$ R2 = 0, otherwise 1-norm != 0
3. Since the multiplication is done on elements on same level of row, it would result the same as computing 1-norms of each then multiplying.
4. In this case 1-norm of $|x + y|$ has to equal $|x|$ + $|y|$ as adding 1-norm after they are added or adding them separately result the same. However we also consider cases where 1-norm of x plus 1-norm of y is bigger than 1-norm of x plus y. It is possible since adding x and y before taking norm does not check the maximum row sum of x and y, which may result smaller 1-norm.

c) Suppose there exist a vector $\in$ R2 that infinity-norm $>$ 2-norm. Infinity-norm is the sum of row, where it takes the maximum element of a vector since it only has two rows. Then 2-norm of this vector is square root of sum of a^2 and b^2, where a and b are the elements in the vector. We supposed infinity-norm is greater than 2-norm however since 2-norm takes square of both elements then square root, which gives a result bigger than one element, the assumption is contradicted. Hence the infinity-norm cannot be larger than 2-norm.

d) Suppose there exist 2-norm of vector x that is greater than 1-norm of x. We are comparing $\sqrt[2]{a^2 + b^2}$ and a+b. Since a+b is always greater than $\sqrt[2]{a^2 + b^2}$ (squaring both formula gives $(a + b)^2$ and $(a^2$+$b^2)$), this assumption is false. Hence 1-norm is always greater 2-norm.

e) We already know properties such as..

λx = Ax
$\|$ λx = Ax $\|$
$|\lambda|$ $\|$x$\|$ = $\|$Ax$\|$ $\leq$ $\|$A$\|$ $\|$x$\|$
From these properties, we can multiply x^{-1} on the right side of x to cancel it out into an identity matrix I. It may result λ = A or may not. From the third property we can say $|\lambda| \leq$ $\|$A$\|$ as it already canceled x out.

f) We need to show an example of $|\lambda| \leq$ 2-norm of matrix A in R^2.
Let A be [[2,4],[5,10]], $\|$A$\|$ equals to 12.04. The eigenvalues for this matrix is 0 and 12, which is smaller than A.

g) P is a projection matrix in R^2, meaning P^2=P. The 1-norm of P is the maximal column sum. We know that P=I-VV^T in which $\|$v$\|$=1 and v^Tv = square of $\|$v$\|$. As norm of v always equal to 1, the norm of vv^T would also equal 1 .