

Homework Assignment hw2

보고서 및 논문 윤리 서약

1. 나는 보고서 및 논문의 내용을 조작하지 않겠습니다.
2. 나는 다른 사람의 보고서 및 논문의 내용을 내 것처럼 무단으로 복사하지 않겠습니다.
3. 나는 다른 사람의 보고서 및 논문의 내용을 참고하거나 인용할 시 참고 및 인용 형식을 갖추고 출처를 반드시 밝히겠습니다.
4. 나는 보고서 및 논문을 대신하여 작성하도록 청탁하지도 청탁받지도 않겠습니다.

나는 보고서 및 논문 작성 시 위법 행위를 하지 않고, 명지인으로서 또한 공학인으로서 나의 양심과 명예를 지킬 것을 약속합니다.



학 과 : 융합소프트웨어학부 데이터테크놀로지전공

과 목 : 인공지능

담당교수 : 전중훈

강좌 번호: 6019

학 번 : 60201901

이 름 : 권성중

(서명)

Artificial Intelligence

Homework Assignment 2.

1. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. (1 pt each)

- a) What is the mean of the data? What is the median? $\text{mean: } 29.962, \text{ median: } 25$
 b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.). bimodal
 c) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data? $Q_1: 20.5, Q_3: 35$
 d) Give the five-number summary of the data. $\text{min: } 13, Q_1: 20.5, Q_2: 25, Q_3: 35, \text{max: } 70$
 e) Draw a boxplot of the data.

- f) How is a quantile-quantile plot different from a quantile plot?

Quantile plot은 1개 값 Quantile, 1개 값은 4개 Q-Q plot은 4개 quantile plot 기리 비교하기 위해 만들어졌다.

2. Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows:

age	frequency
1~5	200
6~15	450
16~20	300
21~50	1500
51~80	700
81~110	44

$$\text{Median} = L + \left(\frac{\frac{n}{2} - \text{freq median}}{\text{freq median}} \right) \text{width}$$

Compute an approximate median value for the data.

3. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

- a) Compute the Euclidean distance between the two objects.
 b) Compute the Manhattan distance between the two objects.
 c) Compute the supremum distance between the two objects.

L	x1	x2	x3	x4
x1	22	20	0	0
x2	1	0	0	0
x3	42	36	0	0
x4	10	8	0	0

L	x1	x2	x3	x4
x1	0	0	0	0
x2	21	0	0	0
x3	20	41	0	0
x4	12	9	32	0

4. Suppose we have the following 2-D data set:

	A ₁	A ₂
x ₁	1.5	1.7
x ₂	2	1.9
x ₃	1.6	1.8
x ₄	1.2	1.5
x ₅	1.5	1.0

	Euclidean	Manhattan	Supremum	Cosine
x ₁	0.14	0.2	0.1	0.99999
x ₂	0.67	0.9	0.6	0.99516
x ₃	0.20	0.4	0.2	0.99999
x ₄	0.22	0.3	0.2	0.99903
x ₅	0.67	0.7	0.6	0.98536

Consider the data as 2-D data points. Given a new data point, $x = (1.4, 1.6)$ as a query, rank the database points based on similarity with the query using Euclidean distance, Manhattan distance, supremum distance, and cosine similarity. $\text{Euclidean: } x_1, x_3, x_4, x_5, x_2, \text{ Manhattan: } x_1, x_4, x_3, x_5, x_2, \text{ Supremum: } x_1, x_3, x_4, x_5, x_2, \text{ Cosine: } x_1, x_3, x_4, x_5, x_2$ (4 pts)

Submitting your assignment :

- Due date: Zip your file and upload it at <https://lms.mju.ac.kr/> by 24:00 Monday March 28th, 2022.
- Your homework cover page must be of the form provided by the <https://lms.mju.ac.kr/>.
- You must zip your homework with the homework cover page, and your homework file name must be of the form "hw2_StudentId_StudentName.zip", i.e., hw2_60063539_김다진.zip.
- You must protect your homework from others. Any form of academic dishonesty will not be tolerated. If you get caught, you will receive -14 points for this homework!
- This assignment is 14 points total and the late penalty is 3 points per day!