# Artificial Intelligence

## Homework Assignment 5.

**1.** Use these methods to normalize the following group of data: (2 pts)

200, 300, 400, 600, 1000

**(a)** min-max normalization by setting *min = 0* and *max = 1*

**(b)** z-score normalization

**2.** Using the data for age and body fat given below, answer the following: (2 pts)

| age | 23 | 23 | 27 | 27 | 39 | 41 | 47 | 49 | 50 |
|---|---|---|---|---|---|---|---|---|---|
| %fat | 9.5 | 26.5 | 7.8 | 17.8 | 31.4 | 25.9 | 27.4 | 27.2 | 31.2 |
| age | 52 | 54 | 54 | 56 | 57 | 58 | 58 | 60 | 61 |
| %fat | 34.6 | 42.5 | 28.8 | 33.4 | 30.2 | 34.1 | 32.9 | 41.2 | 35.7 |

**(a)** Normalize the two attributes based on z-score normalization.

**(b)** Calculate the correlation coefficient (Pearson's product moment coefficient). Are these two attributes positively or negatively correlated? Compute their covariance.

**3.** Consider the following data (in increasing order) for the attribute age: (2 pts)

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

**(a)** Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps.

**(b)** How might you determine outliers in the data? Explain your answer.

**4.** Suppose a group of 12 sales price records has been sorted as follows: (2 pts)

5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215.

Partition them into three bins by each of the following methods:

**(a)** equal-frequency (equal-depth) partitioning

**(b)** equal-width partitioning

**5.** 다음의 데이터를 대상으로 전처리를 수행하고자 한다. A 컬럼은 0~1사이의 값으로 normalization을 수행하고, B 컬럼은 one-hot encoding을 수행한 후 그 결과를 출력하시오. (Python 코드 파일과 실행결과 화면을 캡쳐하여 제출) (2 pts)

| A | B |
|---|---|
| 100 | big |
| 0 | small |
| 40 | medium |
| 80 | big |
| 20 | small |

**Submitting your assignment :**

- Due date: Zip your file and upload it at https://lms.mju.ac.kr by 24:00 Monday April 11th, 2022.
- Your homework cover page must be of the form provided by the lms.
- You must zip your homework with the homework cover page including your jupyter notebook assignment file(*.ipynb) and screen capture image of the execution results of the problem 5. Your homework file name must be of the form "hw5_StudentId_StudentName.zip", i.e., hw5_60063539_장유진.zip.
- You must protect your homework from others. Any form of academic dishonesty will not be tolerated. If you get caught, you will receive -10 points for this homework!
- This assignment is 10 points total and the late penalty is 2 points per day!