# Homework 5

(Advanced) Data Mining: Algorithms and Applications-Winter 2023

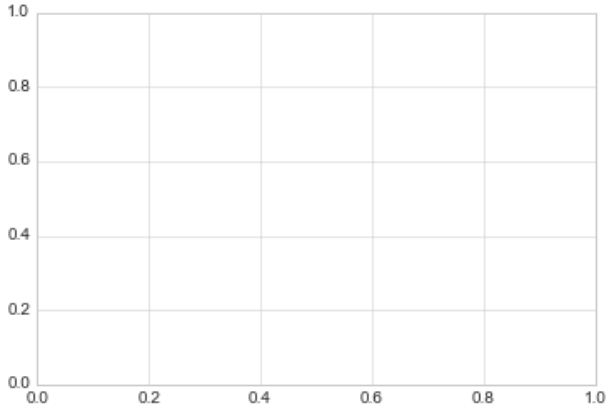
**Due on Apr 15, 11.59PM**

**Important**

- Please type your answers for the calculations.
- Only submissions through Canvas will be accepted.

1. Draw the ROC curve based on the table below and fill the empty columns based on threshold at each step. (20 pts)

| Tuple | # | Class Prob | TP | FN | FP | TN | TPR | FPR |
|---|---|---|---|---|---|---|---|---|
| 1 | p | 0.95 | | | | | | |
| 2 | n | 0.85 | | | | | | |
| 3 | p | 0.78 | | | | | | |
| 4 | p | 0.66 | | | | | | |
| 5 | n | 0.6 | | | | | | |
| 6 | p | 0.55 | | | | | | |
| 7 | n | 0.53 | | | | | | |
| 8 | n | 0.52 | | | | | | |
| 9 | n | 0.51 | | | | | | |
| 10 | p | 0.4 | | | | | | |




| TID | Basket |
|---|---|
| T1 | A, B, E |
| T2 | B, D |
| T3 | B, C |
| T4 | A, B, D |
| T5 | A, C |
| T6 | B, C |
| T7 | A, C |
| T8 | A, B, C, E |
| T9 | A, B, C |

Refer to the table above for questions 2 through 5 and suppose the minimum support is 2.

2. Compute the confidences for the following rules. (10 points)

   (a) $\{A, B\} \Rightarrow E$

   (b) $A \Rightarrow \{B, E\}$

3. Apply the Apriori procedure by using join operations as described on slide (see slide[1]) #15. You need to report all frequent k-itemsets. (20 points)

4. Draw the FP-Tree (see slide #28). (20 points)

5. Import this table by preparing an appropriate input format for Weka and run Apriori algorithm. Please use either .arff or csv format by inspecting sample Weka files. Please report the association rules you find. (15 points)

[1] 06FPBasic_HB_part1_W19.ppt

6. Using Weka, implement Apriori and FP-Growth algorithms on Supermarket data, which is a sample data set coming with Weka installation. You can find it under Weka folder in your system. Please report your results with screen shots. You don't have to report all. Top of the results is enough for this question. (15 points)