10

FIM - Quiz3

January 27, 2025

Name and surname: HOLLÁ PABLO
Instructions

- Each question has one and only one correct answer. Circle the corresponding letter.
- Grading: correct answer (1 pt), incorrect answer (-0.5 pt), no answer (0 pt).

1. Consider the frequent itemset $P = ABCD$ with a frequency of 5. Which of the following statements is correct?
   A. The frequency of the itemset AB must be at least 5.
   B. The frequency of the itemset AB must be at most 5.
   C. The frequency of the itemset AB could be less than 5.
   D. The frequency of the itemset AB must be greater than 5.

2. Consider the association rule $R = (AB \rightarrow C)$ with a confidence of 50%. Which of the following statements is correct?
   A. $\texttt{freq}(C) = 2 \times freq(AB)$.
   B. $\texttt{freq}(AB) = 2 \times freq(C)$.
   C. $\texttt{freq}(C) = 2 \times freq(ABC)$.
   D. $\texttt{freq}(AB) = 2 \times freq(ABC)$.
   E. $\texttt{freq}(C) > freq(AB)$.

   $\hookrightarrow conf(AB \rightarrow C) = \frac{freq(AB \rightarrow C)}{freq(AB)} = \frac{1}{2}$
   $\Leftrightarrow 2\,freq(AB \rightarrow X) = freq(AB)$

3. What is the primary metric used to evaluate the strength of an association rule?
   A. Support
   B. Frequency
   C. Lift
   D. Confidence

4. Which of the following statements about association rule generation is *correct*?
   A. Association rules are generated before finding frequent itemsets.
   B. Rules are generated from frequent itemsets using confidence thresholds.
   C. Only itemsets with exactly two items can generate association rules.
   D. The Apriori property is not relevant for association rule generation.

5. Which algorithm is commonly used to efficiently find frequent itemsets in large datasets?

A. K-Means clustering.

B. The Apriori algorithm.

C. Naïve Bayes classifier.

D. Principal Component Analysis (PCA).

6. Which of the following statements is *not* correct?

A. If an itemset is infrequent, all of its supersets are infrequent.

B. If an itemset is infrequent, all of its subsets are also infrequent.

C. Anti-monotonicity is used to prune the search space in frequent itemset mining.

D. If an itemset is frequent, all of its subsets are also frequent.

7. In the Apriori algorithm, how are candidate itemsets of size k generated?

A. By selecting all possible itemsets of size k from the dataset.

B. By merging itemsets from all previous levels regardless of their frequency.

C. By joining frequents of size $k-1$ and pruning those with infrequent subsets.

D. By directly extracting k-itemsets from the database without iteration.

8. What is the main strategy employed by the Apriori algorithm to reduce candidates?

A. Using depth-first search to explore itemset combinations.

B. Pruning candidate itemsets that contain infrequent subsets.

C. Restricting candidate generation to only 1-itemsets.

D. Generating all possible subsets of the dataset regardless of their frequency.

9. Which of the following heuristics is used by the Apriori algorithm to generate candidate itemsets?

A. Candidate itemsets are generated randomly and then checked for support.

B. Candidate itemsets are generated by merging infrequent itemsets from the previous iteration.

C. Candidate itemsets are generated by joining frequent itemsets from the previous iteration and pruning infrequent ones.

D. Candidate itemsets are generated using a clustering approach to group similar items together.

10. How does the Apriori algorithm use item ordering and prefix trees in candidate generation?

A. Items are ordered randomly to generate candidate itemsets with no particular structure.

B. The algorithm uses lexicographical ordering of items to generate candidate itemsets and a prefix tree to organize them efficiently, with no ordering other than lexicographical being appropriate.

C. The Apriori algorithm orders items by frequency and uses a prefix tree to efficiently generate candidate itemsets and prune infrequent ones.

D. A prefix tree is not used in the Apriori algorithm for candidate generation