| Name: | Roll No: | Marks: |
|---|---|---|

# Data Mining and Machine Learning

## Quiz 1, II Semester, 2025–2026

29 January, 2026

*Some questions below may have more than one correct answer. You get full credit if you select all correct options. You get partial credit if you select a non-empty, strict subset of correct options. You get zero credit if you select any incorrect option.*

1. In the market-basket analysis problem, suppose the set of items $I$ has size $10^7$, the number of transactions $T$ is $10^{10}$ and each transaction $t \in T$ contains at most 10 distinct items. What is the best upper bound we can compute for $F_1$ and $F_2$, the number of frequent itemset of size 1 and 2, respectively, for a support value of 0.01%?

   (a) $F_1 \leq 10^3$ and $F_2 \leq 10^6$.
   (b) $F_1 \leq 10^3$ and $F_2 \leq 4.5 \times 10^4$.
   (c) $F_1 \leq 10^4$ and $F_2 \leq 10^8$.
   (d) $F_1 \leq 10^5$ and $F_2 \leq 4.5 \times 10^5$.

   **Note:** In the original quiz, the last option had a typo and wrongly claimed $F_1 \leq 10^4$.

2. Which of the following strategies can avoid overfitting when building a decision tree.

   (a) Fix an upper bound on the depth of the tree.
   (b) Fix a lower bound on the depth of the tree.
   (c) Fix an upper bound on the size of a leaf node.
   (d) Fix a lower bound on the size of a leaf node.

3. An airport security system consists of a full body scanner followed by manual frisking. If the full body scanner beeps, the passenger is checked manually and then allowed to proceed if there is nothing amiss. If the full body scanner does not beep, no frisking is done. In terms of the entries in the confusion matrix, what ratio should the full body scanner maximize to ensure that no suspicious person is let through unchecked?

   (a) TP/(TP+FP)
   (b) TN/(TN+FP)
   (c) TN/(TN+FN)
   (d) TN/(TN+TP)