[ML] Cold Start Upvote Prediction
Problem ID: coldstart

Note: The following is a fictitious depiction of Quora's cold start ranking system.

Many new answers are added to Quora every day, but the quality of an answer is varying and difficult to determine from the text alone. We often need some user feedback to determine whether an answer is good, but also want to avoid showing a bad answer to many people. In this problem, we will try to determine the expected quality of an answer based only on a small amount of feedback from a small set of "critic" users. Formally, given whether k initial "critic" users would upvote an answer, determine the expected number of total upvotes if the answer was shown to t total users selected from the general population.

For this problem, you will be given a training dataset where both the reactions from the k critic users and the total upvotes after t users see the content are provided, and a testing dataset where only the results of the first k critic users are provided. Your task is to predict the total upvotes after t users see the answers in the testing dataset.

Notes

For this problem we have significantly increased resource limits to allow for ML-based solutions, though you will still need to be mindful of time constraints. Recall that the time limit and memory limit of the evaluation server for each problem is shown in the corresponding problem statement provided on the Challenge website.

Input

Your program will receive input from standard input.

The first line will contain five space-separated integers, k, t, n, m, and u. k is the number of critic reactions provided, t is the number of total users the content will be shown to, n is the number of answers in the training dataset, m is the number of answers in the testing dataset, and u is the number of distinct users in the dataset.

The next $2n$ lines describe the training dataset.

- For the first n lines, the i-th line contains k pairs of integers formatted as follows:

$$\texttt{x\_1:a\_1 x\_2:a\_2 ...x\_k:a\_k}$$

  representing the critic users who saw the i-th answer in the training dataset. a_j is 1 if user x_j saw and upvoted answer i, and 0 if user x_j saw but did not upvote answer i.

- For the next n lines, the i-th line contains a single integer representing the number of upvotes the answer eventually received after being shown to t users.

Finally, m lines follow. The i-th line contains k pairs of integers formatted as follows:

$$\texttt{x\_1:a\_1 x\_2:a\_2 ...x\_k:a\_k}$$

representing the critic users who saw the i-th answer in the testing dataset. a_j is 1 if user x_j saw and upvoted answer i, and 0 if user x_j saw but did not upvote answer i.

Output

Your program should write to standard output.

Print m lines. On the i-th line, print a single integer representing the predicted number of upvotes after t users view the i-th answer in the testing dataset.

Constraints

- $10 \le k \le 20$

- $k < t \le 2 \cdot 10^2$

- $10^3 \le n \le 10^4$

- $0 < m < n$

- $10^2 \le u \le 5 \cdot 10^2$

- $0 \le x_j < u$

Scoring

Score will be proportional to the L2 loss on the test dataset. Baseline solutions were used to generate minimum and maximum L2 losses for scoring.

Sample Explanation

A sample input and output are attached on the challenge website under the statement for this problem. **The sample input is the real first test case**. Note that if your solution is working locally but not in the system, it may be because of TLE.