Day 9: Multiple Linear Regression

HackerRank

Objective
In this challenge, we practice using *multiple linear regression*. Check out the Tutorial tab for learning materials!

Task
Andrea has a simple equation:

$$Y = a + b_1 \cdot f_1 + b_1 \cdot f_2 + \ldots + b_m \cdot f_m$$

for $(m+1)$ real constants (a, f_1, f_2, ..., f_m). We can say that the value of Y depends on m features. Andrea studies this equation for n different feature sets $(f_1, f_2, f_3, \ldots, f_m)$ and records each respective value of Y. If she has q new feature sets, can you help Andrea find the value of Y for each of the sets?

Note: You are not expected to account for bias and variance trade-offs.

Input Format

The first line contains 2 space-separated integers, m (the number of observed features) and n (the number of feature sets Andrea studied), respectively.
Each of the n subsequent lines contain $m+1$ space-separated decimals; the first m elements are features $(f_1, f_2, f_3, \ldots, f_m)$, and the last element is the value of Y for the line's feature set.
The next line contains a single integer, q, denoting the number of feature sets Andrea wants to query for.
Each of the q subsequent lines contains m space-separated decimals describing the feature sets.

Constraints

- $1 \le m \le 10$
- $5 \le n \le 100$
- $0 \le x_i \le 1$
- $0 \le Y \le 10^6$
- $1 \le q \le 100$

Scoring
For each feature set in one test case, we will compute the following:

- $d'_i = \frac{|\texttt{Computed value of Y - Expected value of Y}|}{\texttt{Expected value of Y}}$
- $d_i = \texttt{max(}d'_i\texttt{ - 0.1, 0)}$. We will permit up to a $\pm 10\%$ margin of error.
- $s_i = \texttt{max(1.0 - }d_i\texttt{, 0)}$

The normalized score for each test case will be: $S = \frac{\sum_{i=1}^{q} s_i}{q}$. If the challenge is worth C points, then your score will be $S \times C$.

Output Format

For each of the q feature sets, print the value of Y on a new line (i.e., you must print a total of q lines).

Sample Input

```
2 7
0.18 0.89 109.85
1.0 0.26 155.72
0.92 0.11 137.66
0.07 0.37 76.17
0.85 0.16 139.75
0.99 0.41 162.6
0.87 0.47 151.77
4
0.49 0.18
0.57 0.83
0.56 0.64
0.76 0.18
```

Sample Output

```
105.22
142.68
132.94
129.71
```

Explanation

We're given $m = 2$, so $Y = a + b_1 \cdot f_1 + b_2 \cdot f_2$. We're also given $n = 7$, so we determine that Andrea studied the following feature sets:

- $a + 0.18 \cdot b_1 + 0.89 \cdot b_2 = 109.85$
- $a + 1.0 \cdot b_1 + 0.26 \cdot b_2 = 155.72$
- $a + 0.92 \cdot b_1 + 0.11 \cdot b_2 = 137.66$
- $a + 0.07 \cdot \cdot b_1 + 0.37 \cdot b_2 = 76.17$
- $a + 0.85 \cdot b_1 + 0.16 \cdot b_2 = 139.75$
- $a + 0.99 \cdot b_1 + 0.41 \cdot b_2 = 162.6$
- $a + 0.87 \cdot b_1 + 0.47 \cdot b_2 = 151.77$

We use the information above to find the values of a, b_1, and b_2. Then, we find the value of Y for each of the q feature sets.