Polynomial Regression: Office Prices

For those new to the topic of Polynomial Regression, here's a wonderful video of Dr. Andrew Ng from Stanford, explaining the topic. It might be helpful to watch this video and any other resources of your choice before proceeding with the problem.

Polynomial Regression

The Problem

Charlie wants to purchase office-space. He does a detailed survey of the offices and corporate complexes in the area, and tries to quantify a lot of factors, such as the distance of the offices from residential and other commercial areas, schools and workplaces; the reputation of the construction companies and builders involved in constructing the apartments; the distance of the offices from highways, freeways and important roads; the facilities around the office space and so on.

Each of these factors are quantified, normalized and mapped to values on a scale of 0 to 1. Charlie then makes a table. Each row in the table corresponds to Charlie's observations for a particular house. If Charlie has observed and noted F features, the row contains F values separated by a single space, followed by the *office-space price in dollars/square-foot*. If Charlie makes observations for H houses, his observation table has (F+1) columns and H rows, and a total of (F+1) * H entries.

Charlie does several such surveys and provides you with the tabulated data. At the end of these tables are some rows which have just F columns (the price per square foot is missing). Your task is to predict these prices. F can be any integer number between 1 and 5, both inclusive.

There is one important observation which Charlie has made.

- The prices per square foot, are (approximately) a polynomial function of the features in the observation table. This polynomial always has an order less than 4

Input Format

The first line contains two space separated integers, F and N. Over here, F is the number of observed features. N is the number of rows for which features as well as price per square-foot have been noted. This is followed by a table having F+1 columns and N rows with each row in a new line and each column separated by a single space. The last column is the price per square foot.

The table is immediately followed by integer T followed by T rows containing F columns.

Constraints

1 <= F <= 5
5 <= N <= 100
1 <= T <= 100
0 <= Price Per Square Foot <= 10^6 0 <= Factor Values <= 1

Output Format

T lines. Each line 'i' contains the predicted price for the 'i'th test case.

Sample Input

```
2 100
0.44 0.68 511.14
0.99 0.23 717.1
0.84 0.29 607.91
0.28 0.45 270.4
0.07 0.83 289.88
0.66 0.8 830.85
0.73 0.92 1038.09
0.57 0.43 455.19
0.43 0.89 640.17
0.27 0.95 511.06
0.43 0.06 177.03
0.87 0.91 1242.52
0.78 0.69 891.37
0.9 0.94 1339.72
0.41 0.06 169.88
0.52 0.17 276.05
0.47 0.66 517.43
0.65 0.43 522.25
0.85 0.64 932.21
0.93 0.44 851.25
0.41 0.93 640.11
0.36 0.43 308.68
0.78 0.85 1046.05
0.69 0.07 332.4
0.04 0.52 171.85
0.17 0.15 109.55
0.68 0.13 361.97
0.84 0.6 872.21
0.38 0.4 303.7
0.12 0.65 256.38
0.62 0.17 341.2
0.79 0.97 1194.63
0.82 0.04 408.6
0.91 0.53 895.54
0.35 0.85 518.25
0.57 0.69 638.75
0.52 0.22 301.9
0.31 0.15 163.38
0.6 0.02 240.77
0.99 0.91 1449.05
0.48 0.76 609.0
0.3 0.19 174.59
0.58 0.62 593.45
0.65 0.17 355.96
0.6 0.69 671.46
0.95 0.76 1193.7
0.47 0.23 278.88
0.15 0.96 411.4
0.01 0.03 42.08
0.26 0.23 166.19
0.01 0.11 58.62
0.45 0.87 642.45
0.09 0.97 368.14
0.96 0.25 702.78
0.63 0.58 615.74
0.06 0.42 143.79
0.1 0.24 109.0
0.26 0.62 328.28
0.41 0.15 205.16
0.91 0.95 1360.49
0.83 0.64 905.83
0.44 0.64 487.33
0.2 0.4 202.76
0.43 0.12 202.01
0.21 0.22 148.87
0.88 0.4 745.3
```

```
0.31 0.87 503.04
0.99 0.99 1563.82
0.23 0.26 165.21
0.79 0.12 438.4
0.02 0.28 98.47
0.89 0.48 819.63
0.02 0.56 174.44
0.92 0.03 483.13
0.72 0.34 534.24
0.3 0.99 572.31
0.86 0.66 957.61
0.47 0.65 518.29
0.79 0.94 1143.49
0.82 0.96 1211.31
0.9 0.42 784.74
0.19 0.62 283.7
0.7 0.57 684.38
0.7 0.61 719.46
0.69 0.0 292.23
0.98 0.3 775.68
0.3 0.08 130.77
0.85 0.49 801.6
0.73 0.01 323.55
1.0 0.23 726.9
0.42 0.94 661.12
0.49 0.98 771.11
0.89 0.68 1016.14
0.22 0.46 237.69
0.34 0.5 325.89
0.99 0.13 636.22
0.28 0.46 272.12
0.87 0.36 696.65
0.23 0.87 434.53
0.77 0.36 593.86
4
0.05 0.54
0.91 0.91
0.31 0.76
0.51 0.31
```

Sample Output

```
180.38
1312.07
440.13
343.72
```

Explanation

There are two features which have been noted by Charlie. There are 100 data points, for which he has taken note of the features, and the price per square foot (in the last column).

At the end, are four rows where he knows the two features, you output the predicted price/square foot of the office space for every testcase.

Recommended Technique

Use a regression based technique. At this point, you are not expected to account for bias and variance trade-offs.

Scoring

For each test in a test case file we will compute the following:

```
d = Normalized Distance from expected answer
  = abs(Computed-Expected)/Expected
```

Since there can be multiple ways to approach this problem, which account for bias, variance, various subjective factors and "noise", we will take a realistic approach to scoring, and permit upto +/- 10% swing of our expected answer.

```
d_adjusted = max(d - 0.1, 0)
Score for each test = max(1-d_adjusted,0)
Score for the test case file == (Average of the scores for the tests it contains) * M
```

Where M is the Max possible score for the test case.

Suppose we have a test case file with just one test. Suppose our expected answer is 10
And your answer is: 9.5
d = (10 - 9.5) / 10 = 0.05
d_adjusted = max(0.05 - 0.1,0) = 0 Score = max(1-d_adjusted, 0) = max(1,0) = 1