

# Predict the Missing Grade

## Introduction

The [CBSE](#) Class 12 examination, is taken by Indian high school students at the end of K-12 school education. The scores or grades in this examination form the basis of their entry to the College or University system, for an undergraduate program. At the K-12 level, students appear for examination in five subjects. These five subjects generally include one language; three elective subjects oriented towards Science, Commerce or Humanities; and any elective of their choice as a fifth subject.



## The Challenge

This challenge is based on real school data of the CBSE Class 12 examination conducted in the year 2013. You are given the grades obtained by students with specific but popular combinations of subjects (and all these students had opted for Mathematics). Their grades in four subjects are known to you. However their grade in Mathematics (i.e, the fifth subject) is hidden.

The records provided to you are the grades obtained by students who had opted for the following combinations of subjects or courses **and obtained a passing grade in each subject**. The individual subjects in the data are:

English, Physics, Chemistry, Mathematics, Computer Science, Biology, Physical Education, Economics, Accountancy and Business Studies.

The most dominant subject combinations, account for approximately 99% of the data are:

English, Physics, Chemistry, Mathematics, Computer Science  
English, Physics, Chemistry, Mathematics, Physical Education  
English, Physics, Chemistry, Mathematics, Economics  
English, Physics, Chemistry, Mathematics, Biology  
English, Economics, Accountancy, Mathematics, Business Studies

The grades of students in four subjects (other than Mathematics) are provided to you. Can you predict what grade they had obtained in Mathematics?

To help you build a prediction engine, we will provide you with a training file, containing the grade points obtained by students with the above subject combinations, in all five subjects.

## Notes about the Grading System

The student is first assessed on a scale of 100. (S)He needs a score of at least 33% to pass in the subject. Among those who pass:

Grade 1 is assigned to the top one-eighth of students who pass the course.  
Grade 2 is assigned to the next one-eighth of students who pass the course.  
.....  
Grade 8 is assigned to the last one-eighth of students who pass the course.

If more than 1 student share the same score and lie in the margin, they share the higher grade.

## Input Format

The first line will be an integer N. N lines follow each line being a valid [JSON](#) object. The following fields of raw data are given in json.

SerialNumber (Numeric): The identifier of the student record. This is provided just for identification purposes and does not have any direct use.  
English (numeric) : The grade (between 1 and 8) obtained in English. This will always be present.  
Three more numeric fields from among: Physics, Chemistry, ComputerScience, Hindi, Biology, PhysicalEducation, Economics, Accountancy, BusinessStudies.

The input for each record has the grade for all subjects opted by a student, other than *Mathematics* which you have to predict as the answer.

## Constraints

$$1 \leq N \leq 10^5$$

The SerialNumber field will contain a unique numeric identifier such that  $1 \leq \text{SerialNumber} \leq 5 * 10^5$ .

All other fields in the JSON fragment will represent the grades obtained in four subjects and will be populated by numeric values between 1 and 8, both inclusive.

## Output Format

For each student record that is given as a JSON object, containing the grade obtained in four subjects, output the predicted grade in Mathematics (this will be a numeral between 1 and 8, both inclusive) in a newline.

## Training File and Sample Tests

The [training file with sample test data](#) is available here. The training file is also present in the current directory in which your code is executed by the name of "training.json".

The three files in this package are:

training.json  
sample-test.in.json  
sample-test.out.json

Training data as well as sample testcases have been provided in the above file for offline training and to help you build your prediction model. When you submit your solution to us, you can assume that the training file can be accessed by reading "training.json" which will be placed in the same folder as the one in which your program is being executed.

## Sample Input

```
12345
{"SerialNumber":1,"English":1,"Physics":2,"Chemistry":3,"ComputerScience":2}
json_object
json_object
json_object
.
.
.
json_object
```

## Sample Output

```
1
3
4
7
8
....
....
....
```

## Explanation

It is predicted that first candidate obtained grade 1 in Mathematics, the second candidate achieved grade 3 in Mathematics, the third candidate achieved grade 4 in Mathematics and so on.

## Scoring

For each of the N records in the input file, we will compute:

$$p = \text{abs}(\text{Predicted Grade Point in Mathematics} - \text{Actual Grade Point in Mathematics})$$

Where 'abs' indicates the Absolute Value or Magnitude. If  $p = 0$  or  $1$  your answer for that particular student record will be considered correct. i.e, we allow a tolerance of one grade point away from the correct answer, to take into consideration the marginal errors which might occur during the testing or grading process.

$$\text{Score} = 100 * ((C-W)/N)$$

Where C = Number of Correct predictions, not more than one grade point away from the actual grade point assigned.

W = Number of wrong (incorrect) predictions and

N = Total number of records in the input.

While the contest is in progress, only the score based on the sample test case will be displayed to you. After the contest is completed, we will revise the scores based on performance on a hidden test set only.

However, when you make submissions, you will be able to see whether your program attains a positive score on both the sample and the hidden test cases (to avoid a situation where unexpected errors occur on the hidden test set at the end).