C. Blood Pressure Estimation

time limit per test: 1 second
memory limit per test: 256 megabytes
input: standard input
output: standard output

The task is to estimate systolic (SBP) and diastolic (DBP) blood pressure for a given individual by recorded photoplethysmography (PPG) and electrocardiography (ECG) signals. For every individual there are several known "calibration" records. That is, about 20% of all records that contains SBP and DBP values along with the corresponding PPG and ECG signal traces is given for each person.

Background

Blood pressure is pressure that blood applies to the wall of a particular artery (with respect to atmospheric pressure). Since the flux of blood is pulsating, blood pressure is usually characterized by two values: minimal (diastolic) and maximal (systolic) blood pressure. Under various conditions blood pressure for a given individual may be different. Measuring of blood pressure usually requires a 1 minute procedure during which an artery in the hand is compressed by a cuff. Since the procedure is relatively sophisticated, simpler methods of estimation of blood pressure are potentially of great value.

Photoplethysmography (PPG) is an optical method for detection changes in blood volume in a part of a body (in our case it is a finger). A PPG signal corresponds to the amount of light absorbed by the finger, and thus depends on the amount of blood in the finger. Electrocardiography (ECG) is a method of registering electrical activity of the heart. An ECG signal corresponds to the difference in electrical potential of two electrodes located on the body.

Both ECG and PPG signals implicitly characterize the state of the cardiovascular system of the body. Therefore, it seems reasonable that these signals can be used to estimate blood pressure. However, the problem to predict blood pressure for all population using only the ECG and PPG signals is quite difficult. For this competition, a weaker problem is posed: predict SBP and DBP for each given individual under different conditions provided that several "calibration" records for that individual are known. The calibration records contain the values of SBP and DBP as well as the traces of PPG and ECG signals.

Input

You can download the data by the link https://assets.codeforces.com/files/huawei/blood-pressure-estimation-data.zip.

The data files are split across 3 folders: `data_train`, `data_test1_blank`, and `data_test2_blank`. Each file name has the following form:

subjXXlogYYYY.csv
where XX is the identity number of a person and YYYY is the record number. Each file has the following contents:

SBP,DBP
PPG_1,ECG_1
PPG_2,ECG_2
PPG_3,ECG_3
....,....
where the first row contains the values of SBP and DBP, and the subsequent rows contain the samples of the PPG and ECG signals. The sampling rate is 500 Hz.

Please note that parts of some ECG and PPG signal traces may be corrupted. For instance, it can happen that the ECG electrodes are applied only after the start of recording. This leads to the initial part of the ECG signal trace being absolutely not valid. Another problem that sometimes happen is saturation of PPG signal. Be careful to preprocess the broken traces first.

The files located in `data_train` folder are intended for training. All such files contain the ground truth SBP a DBP values.

The files located in `data_test1_blank` folder are intended for preliminary submissions. For each identity number XX there are about 20% of files `subjXXlogYYYY.csv` containing the ground truth values SBP and DBP; these are the "calibration" records. The other files in `data_test1_blank` folder contain `0,0` instead of `SBP,DBP` as their first row.

Output

For making a submission you need to generate a CSV file with the following contents:

subjXXlogYYYY.csv,SBP1,DBP1
subjXXlogYYYY.csv,SBP2,DBP2
......,....,....
where the first column is the file name from the `data_test1_blank` folder, and the last two columns contain your estimation for SBP and DBP (integers between 0 and 1000). You need to include the files with the "calibration" record. The score of your submission is determined as a weighted sum of root mean squared error (RMSE) of SBP and DBP, with the error in DBP having 2 times larger weight than the error in SBP:

$$Score = 100 \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{SBP}_i - \mathrm{SBP}_{i,\mathrm{gt}} \right)^2} + 200 \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{DBP}_i - \mathrm{DBP}_{i,\mathrm{gt}} \right)^2},$$

where $\mathrm{SBP}_{\bullet,\mathrm{gt}}$ and $\mathrm{DBP}_{\bullet,\mathrm{gt}}$ are the ground truth blood pressure values, and N is the total number of records including the "calibration" records.

The files located in `data_test2_blank` folder are intended for the final submission. As well as for the `data_test1_blank` folder, the ground truth values of SBP and DBP are only given for 20% of all files.

Judging

The problem A contains the tests from the `data_train` folder. The answers for the training data are open but you may submit your answers if you want. This problem is only for testing your solutions. This problem will be hidden before the official system testing.

The problem B contains the tests from the `data_test1_blank` folder. Your solutions will be judged and scored immediately after each submission. Use this problem to estimate the quality of your approach. This problem will be hidden before the official system testing.

The problem C contains the tests from the `data_test2_blank` folder (i.e. the official system tests) which will be used to determine the winners. Your submissions will not be judged during the contest, however the format of submissions will still be checked. Any well-formed submission will receive 0 points during the contest. After the contest for each participant we will use the *latest* well-formed (0 points) submission on the problem C to rejudge it and calculate the scores. These are the scores that will be used to determine the winners.