Computer Science 130B
Spring 2017
Programming Assignment #4

Due: 11:59pm, Wednesday June 7

You are given a number of sample points (around 100-1000) along the flight path of a missile. Unfortunately the samples are noisy and may contain an unknown number of outliers (less than 30% of the samples for sure). You are to implement a randomized algorithm to obtain a best estimate of the missile's flight path. To simplify the problem, we assume that the observed flight path is a straight line of the form y=ax+b and lies in a plane (the flight plane).

The input sample points will be given to you on the stdin as *n* 2D coordinates in the flight plane (x_i, y_i), one per line. You are to output the coefficients (a, b) of the best fitting line equation y=ax+b on the stdout.

Again, the randomized algorithm should look like the following:

- Repeat a number of times
    - Select seed points (2) randomly from the input
    - Compute the line equation based on the chosen seed points
    - Calculate the distances of the remaining n-2 points to the best fitting line and record the median error, and
    - Retain the current trial solution if the median error is less than that of all previous trials
- For the retained best solution from the randomized trials above
    - Perform a least-square line fit using all sample points with a fitting error less than the median (i.e., the best half of the sample). Output the fitting coefficient (a,b) as the best estimate.

You should design your algorithm to fill in all missing details - your algorithm should handle noise and outliers and should be reasonably efficient. Some sample results are shown below (red +: input points, blue line: the best fitting line)