

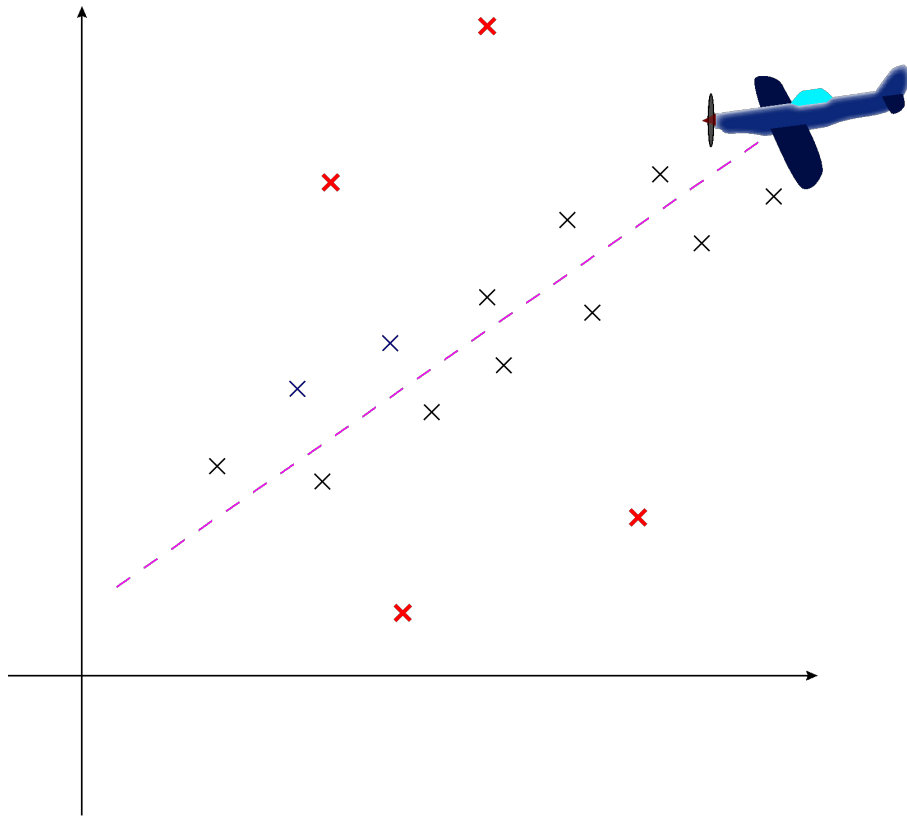
CS 130B—Data Structures and Algorithms II

Discussion Section Week 9

Programming Assignment 4

- Due June 7th Wednesday at 11:59pm

Programming Assignment 4



We need to estimate the airplane's trajectory or in the presence of noise

Naïve Approach

- If ϵ is the fraction of outlier data points, then $n(1 - \epsilon)$ points are usable
- Try all $\binom{n}{n(1 - \epsilon)}$ combinations of points and find the one with the closest fit

Tractability Problem

- If you have 1000 points, as in the problem, and 30% are outliers,

$$\binom{1000}{700} \propto 10^{263}$$

How big is 10^{263} ?

According to the standard model, there are about 10^{80} particles in the known universe, and 10^{17} seconds have gone by since the beginning of time..

Even if every particle in the universe was a 3 GHz computer running since the big bang, by now we would have performed $10^{17} \times 10^9 \times 10^{80} = 10^{106}$ operations. 10^{263} is very thoroughly unobtainable.

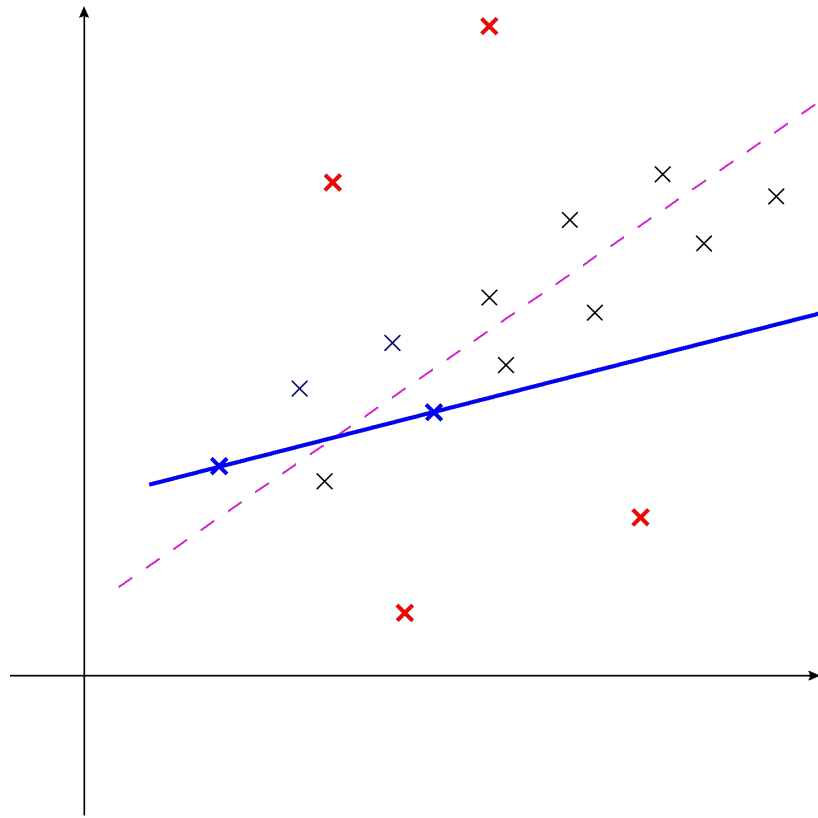
Could we Pick them Randomly?

The number of trial you need to run to have a T percentage change of success is,

$$m = \frac{\ln(1 - T)}{\ln(1 - (1 - \epsilon)^p)}$$

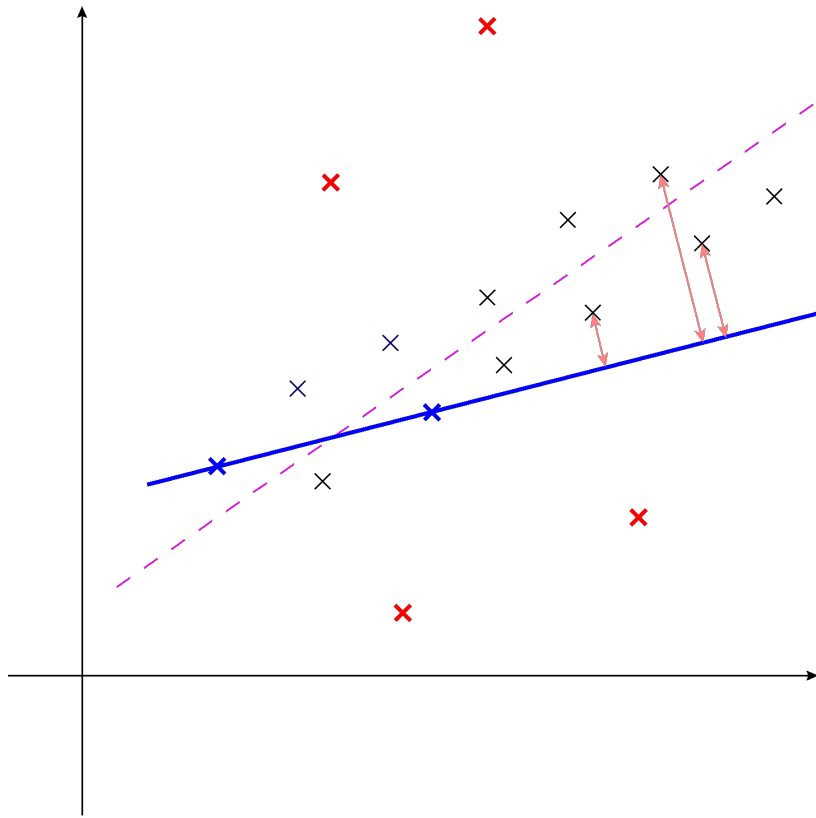
For $p = 1000$ and $T = 90\%$, m is larger than 10^{120} trials..

Randomized Median Method



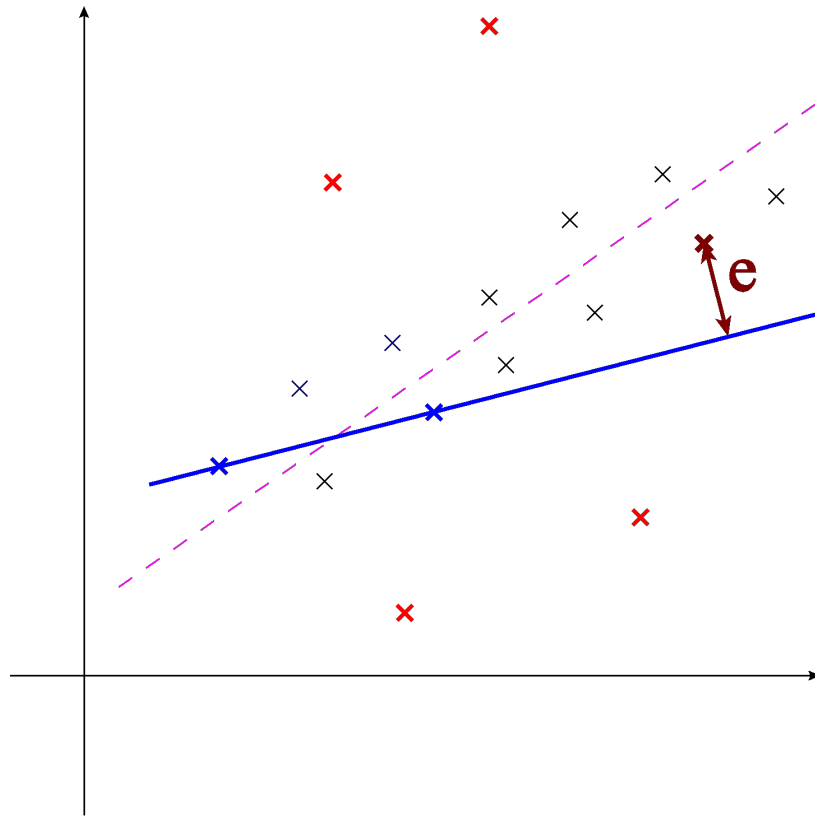
Pick two points and find the equation for the line passing through the points

Errors



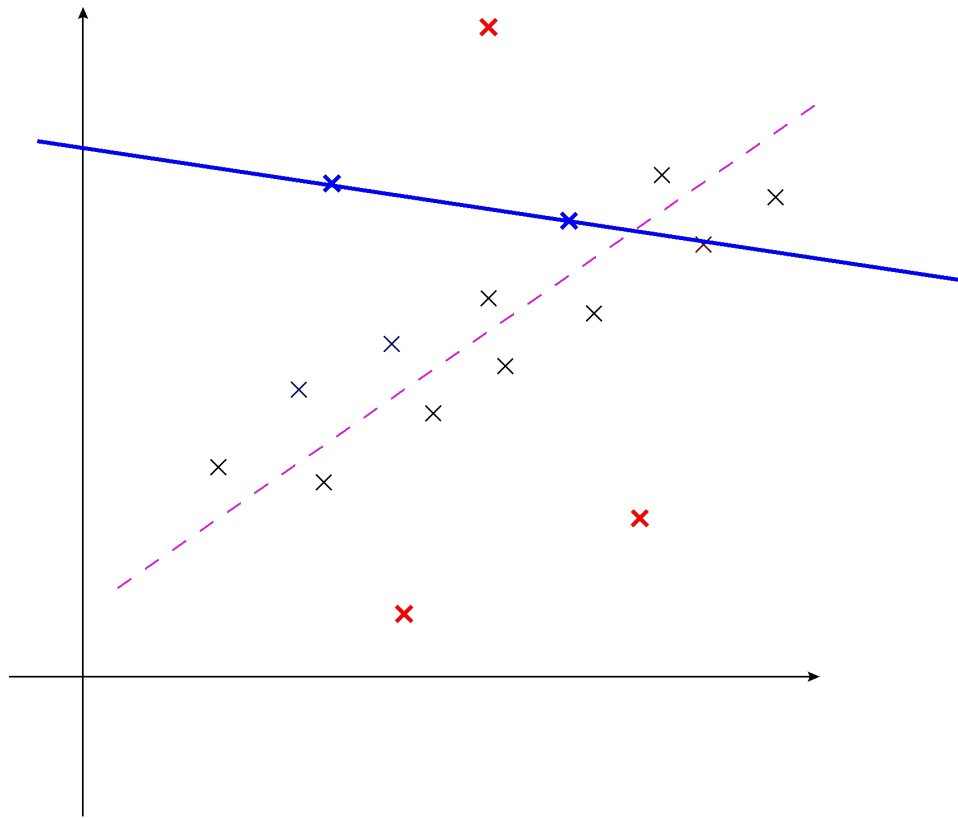
Compute the distances of each point to the line

Find the Median Error



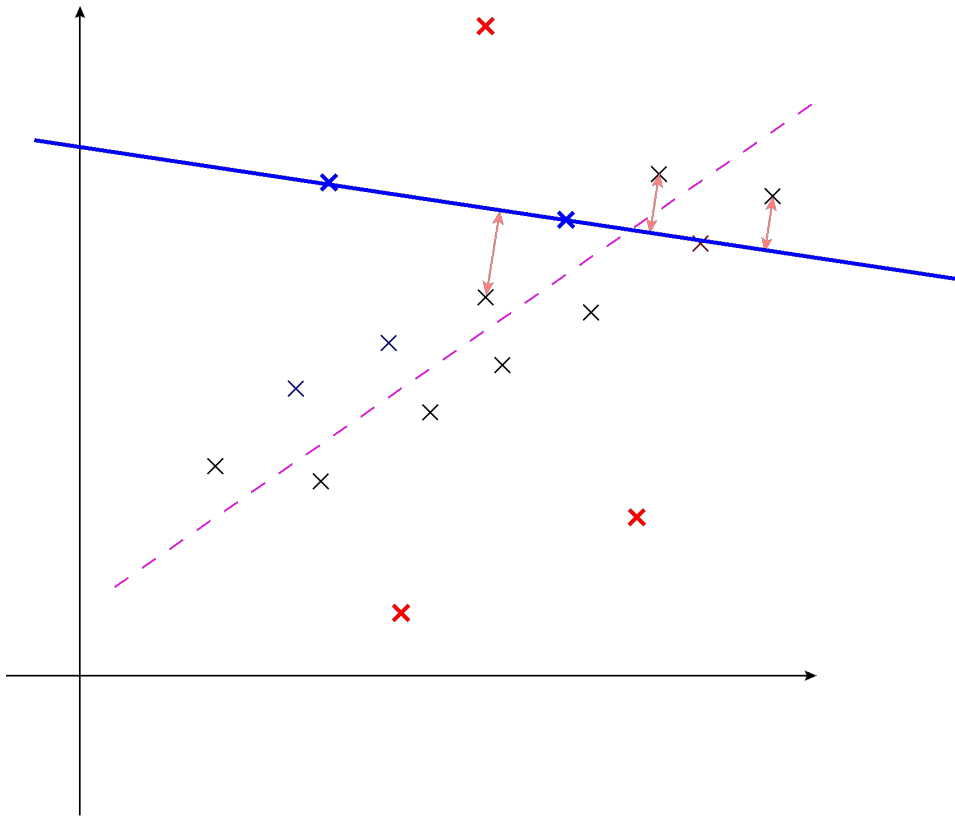
Store the median error in memory

Two Other Points



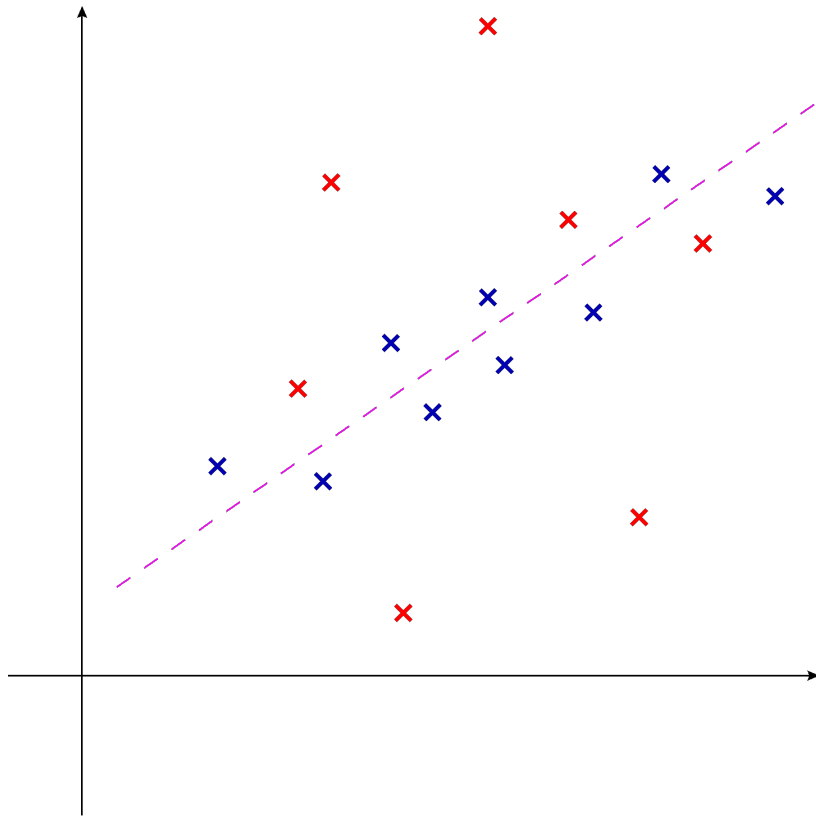
Pick two different points

Errors



Compute the new errors and store error median

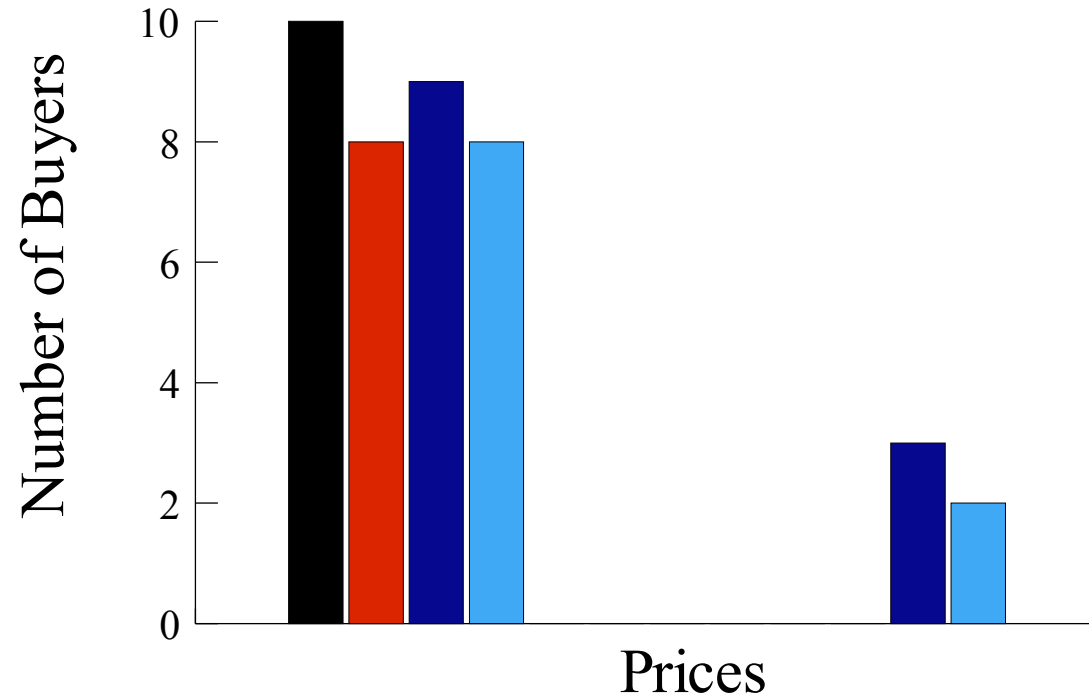
Final Step



Find the two points that resulted in the smallest median error. Pick half of the points with the smallest error. Recompute the line as a least-squares fit with this set of points.

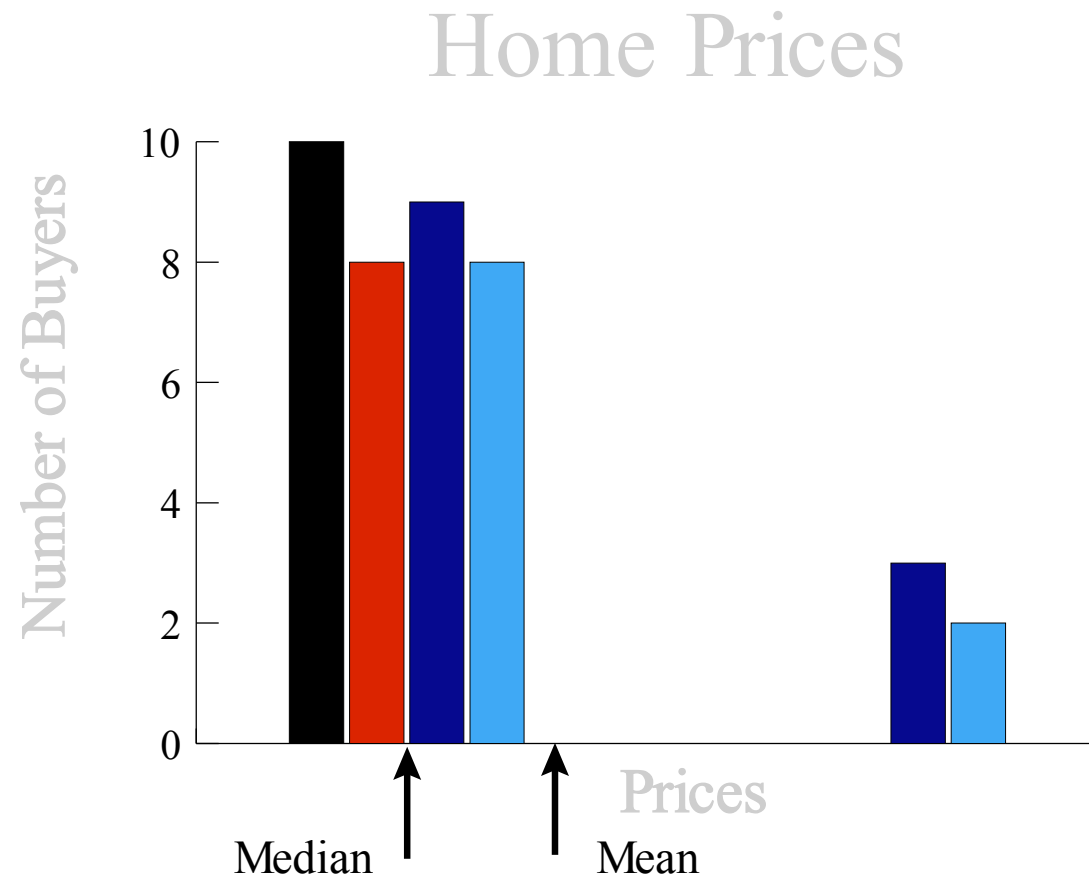
Why/How it works

Home Prices



The median is a better indicator of a “typical value” when the data has outliers.

Why/How it works



The median is a better indicator of a “typical value” when the data has outliers.