Ordinary least squares

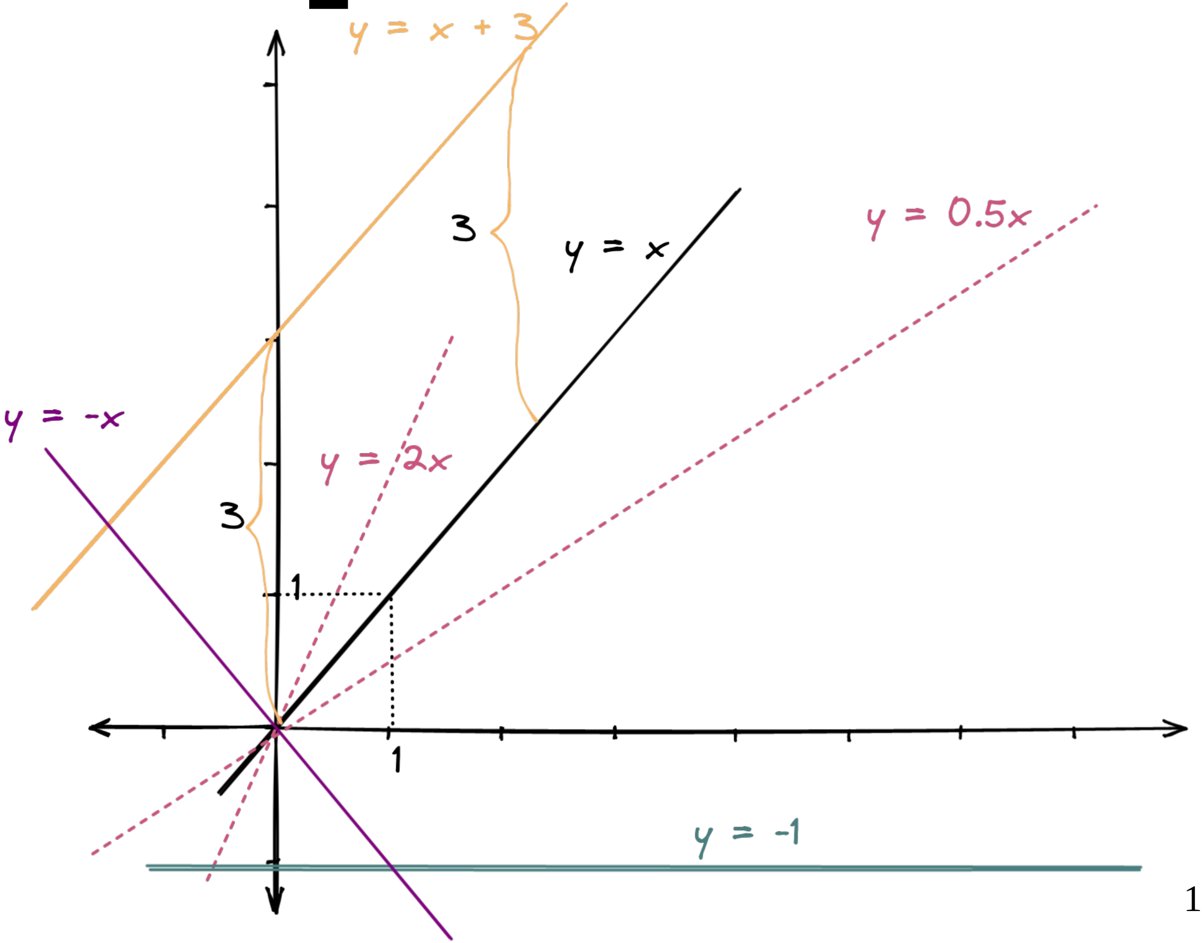
Any line can be described by the following equation:

y = ax + b

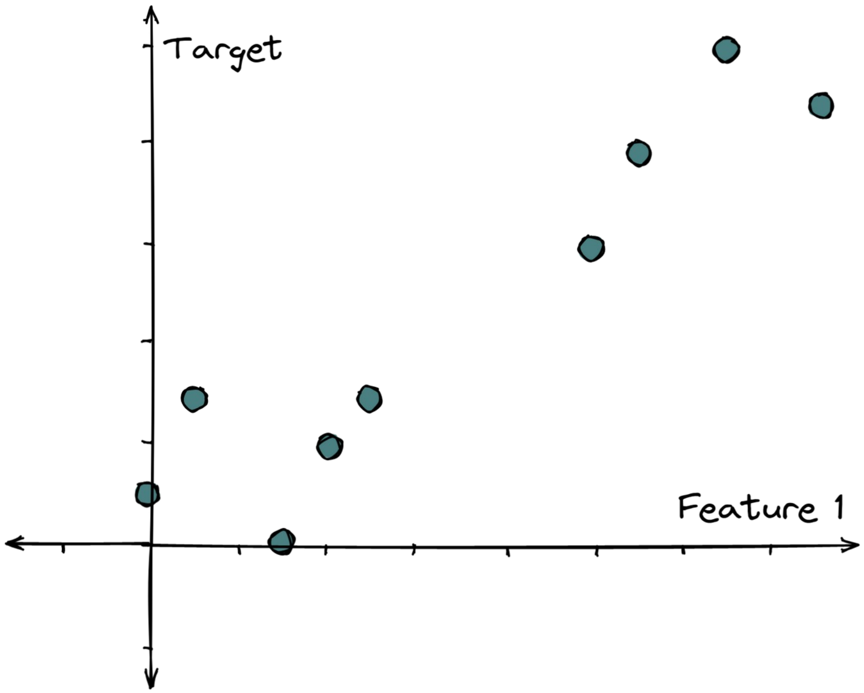
a – slope
b – intercept

Slope (a) is the change in **y** compared to the change in **x.**

Intercept (b) describe the shift of the equation relatively to y-axis

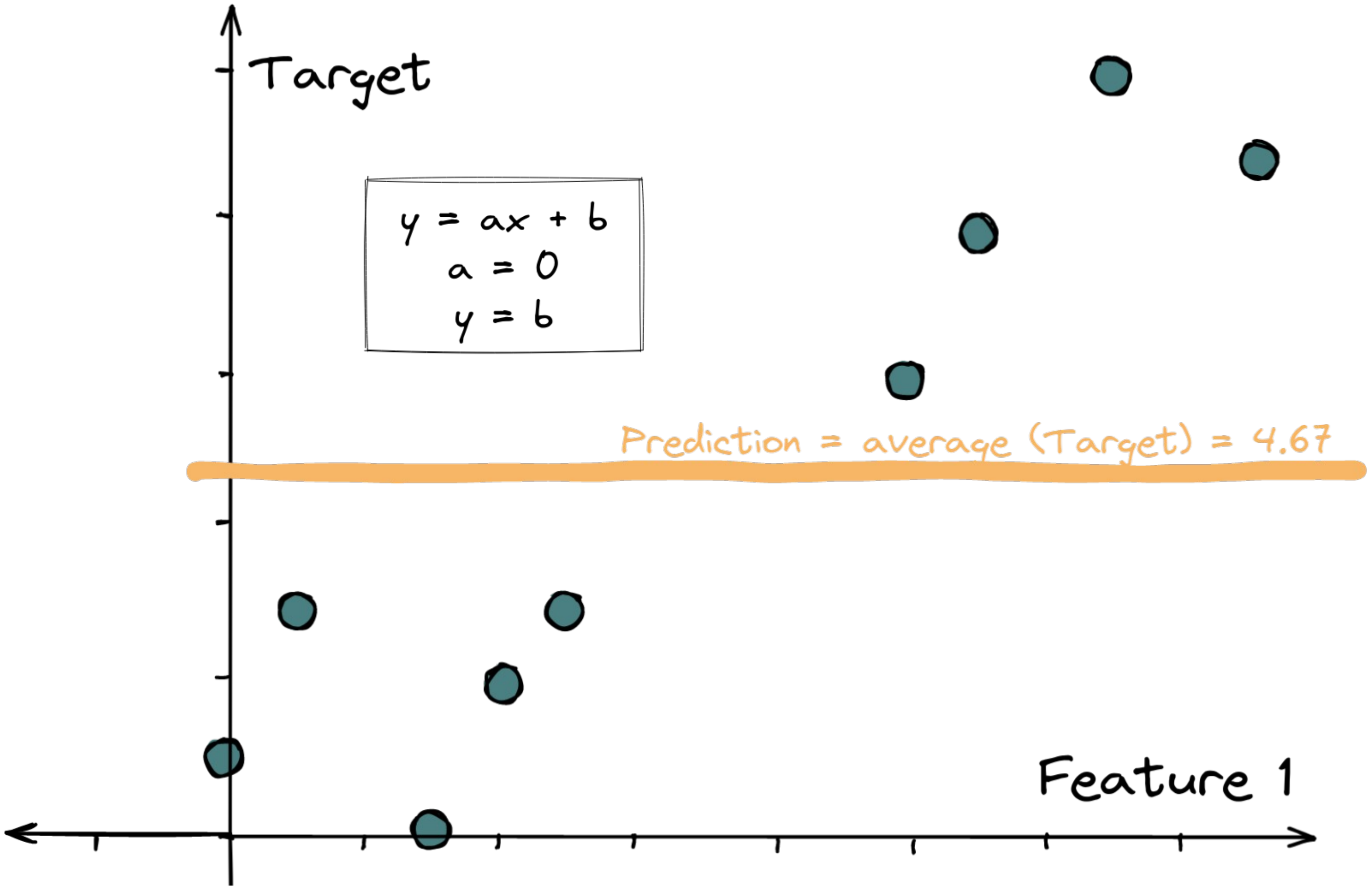


OLS. Step1 : Plot Data



Feature 1	Target
0	1
2	3
6	0
8	2
10	3
20	6
22	8
26	10
30	9

OLS. Step2 : Find mean



OLS. Step3a : Find errors

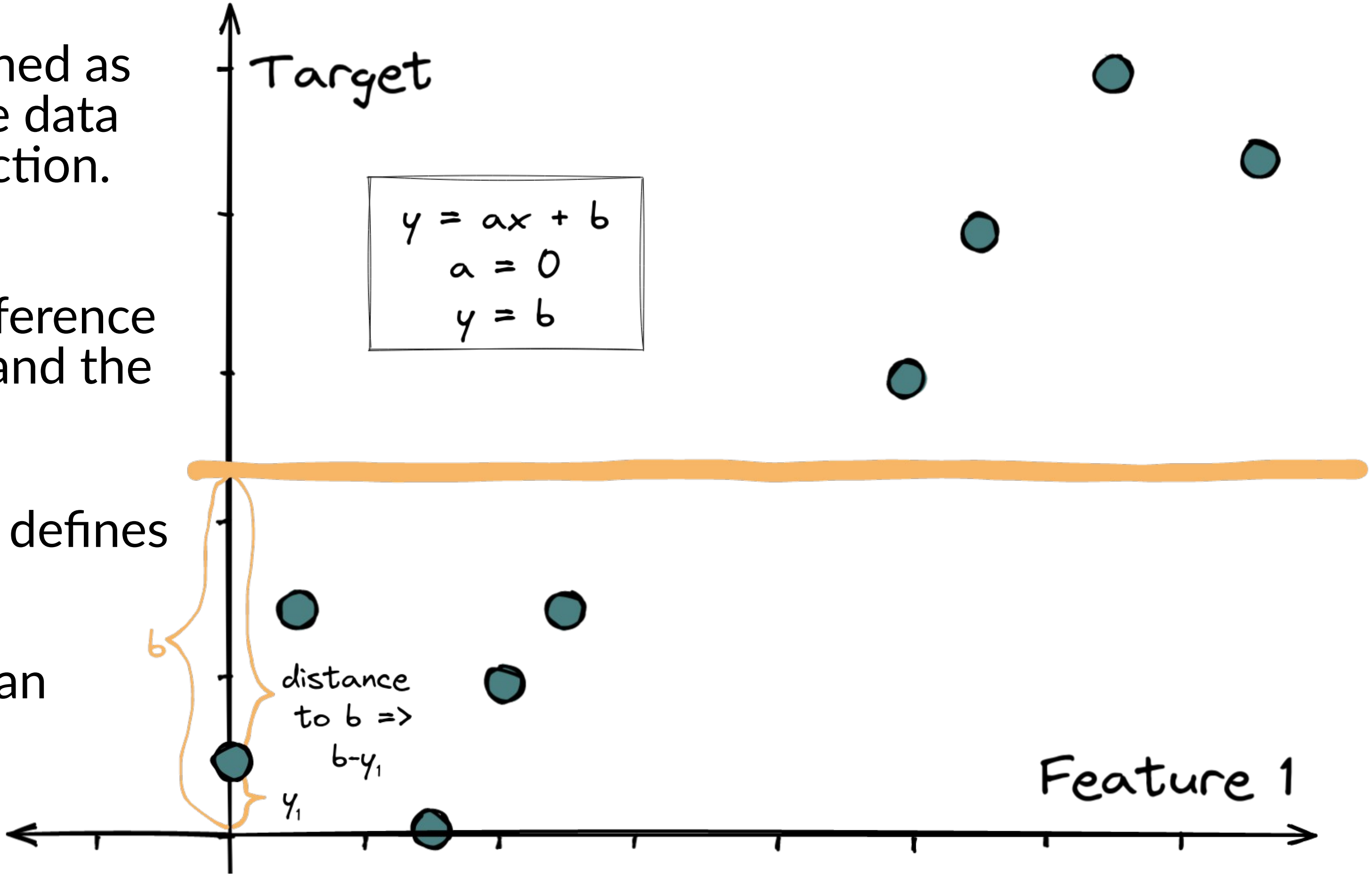
Errors can be defined as distances from the data point to the prediction.

In this case, we're calculating the difference between value y_1 and the mean for all data.

The equation that defines the mean is

y = b,

where b is the mean



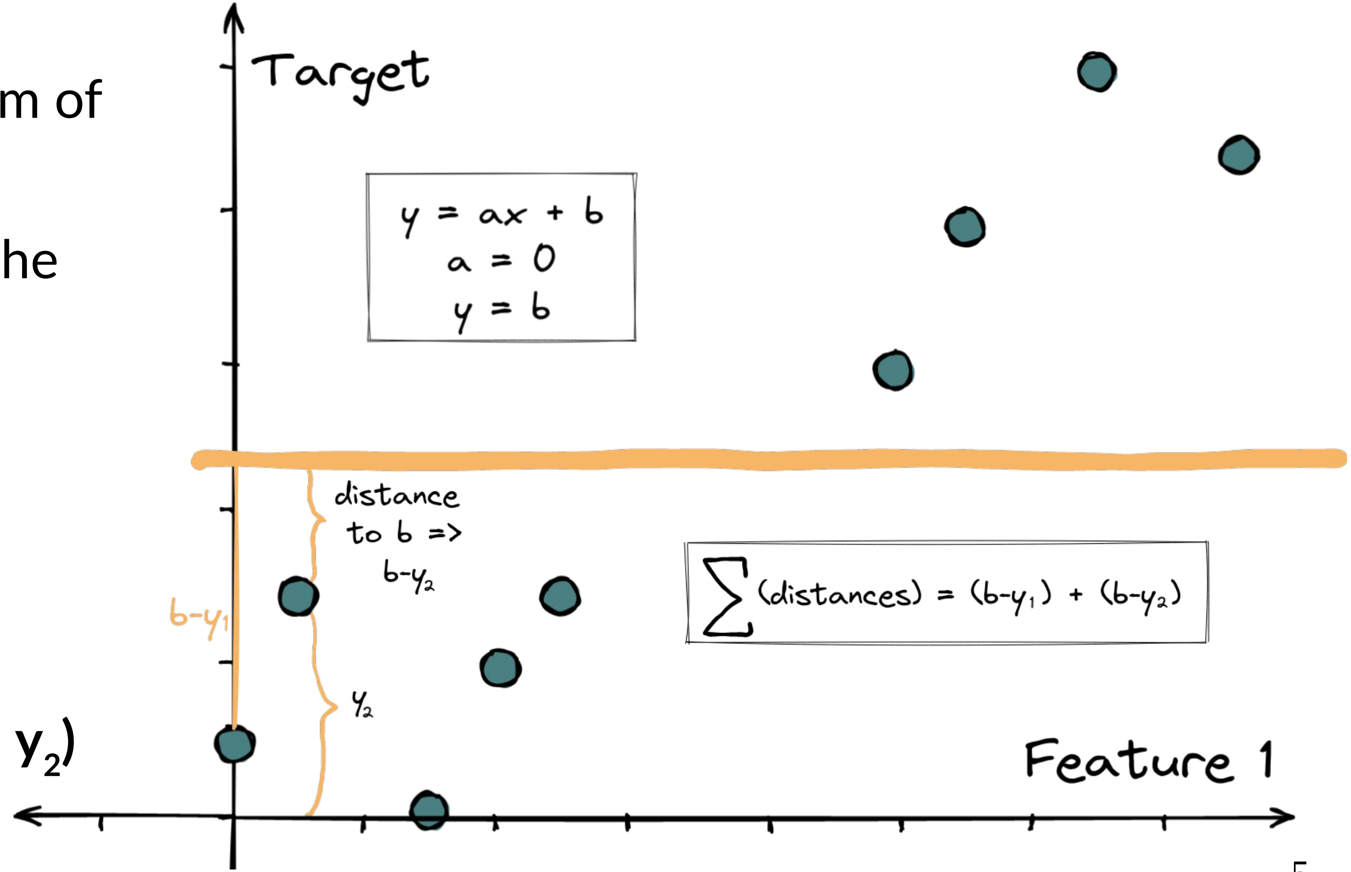
OLS. Step3b : Find errors

We start calculating the sum of errors for all data points.

The idea is that the lower the sum, the better is our prediction.

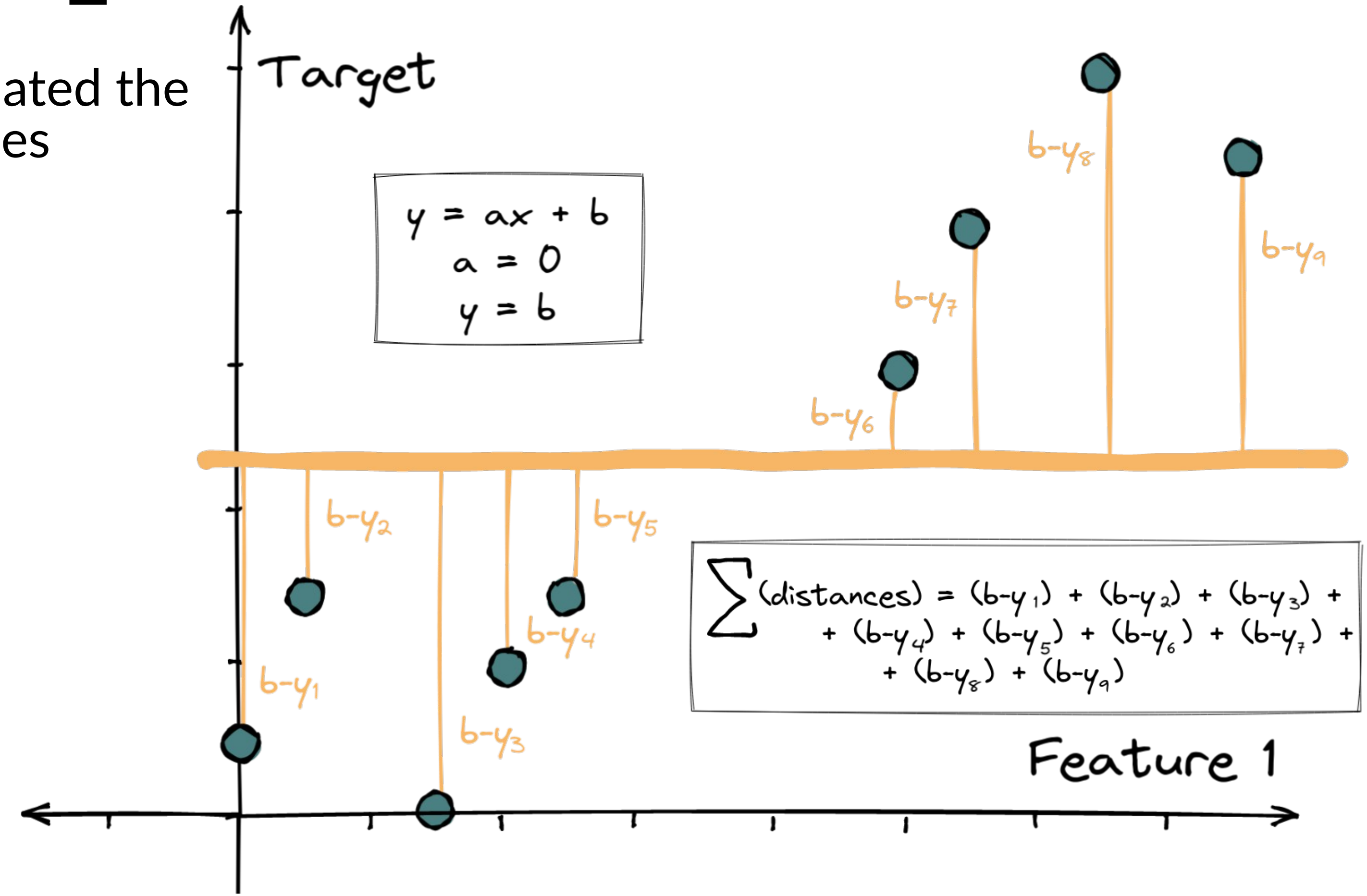
For data point 1: **b – y_1**
For data point 2: **b – y_2**

Current sum: (b – y_1) + (b – y_2)



OLS. Step3c : Find errors

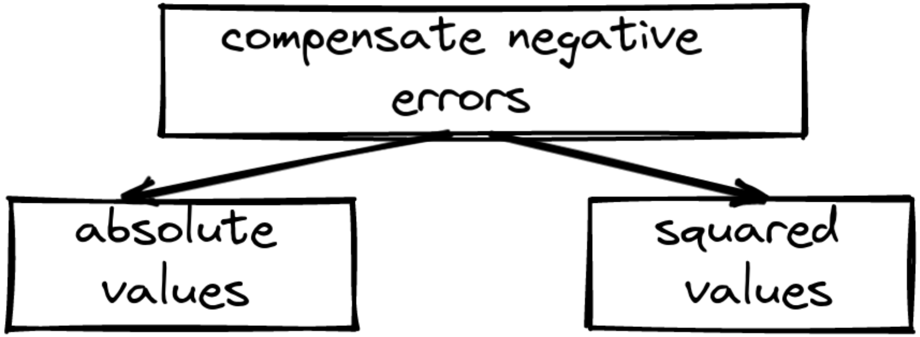
Now we’ve calculated the sum of all distances



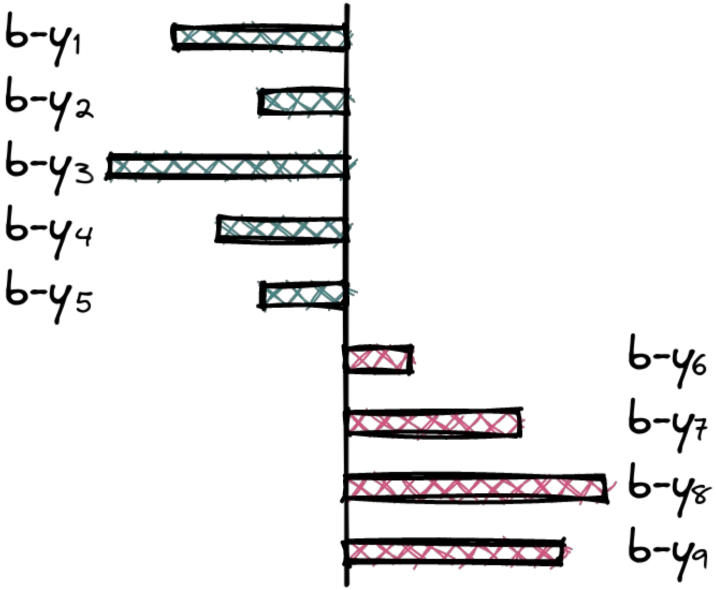
OLS. Step4 : Analyze errors

The idea of calculating the sum of errors is good. But, as you can see, negative errors compensate positive error values. The total sum, in our case, is 0.

To overcome this problem, we can either calculate absolute values or squared values. Both methods will result in compensating negative errors.



$$\sum(\text{distances}) = (b-y_1) + (b-y_2) + (b-y_3) \\ (b-y_4) + (b-y_5) + (b-y_6) + (b-y_7) \\ + (b-y_8) + (b-y_9) \\ = 0$$

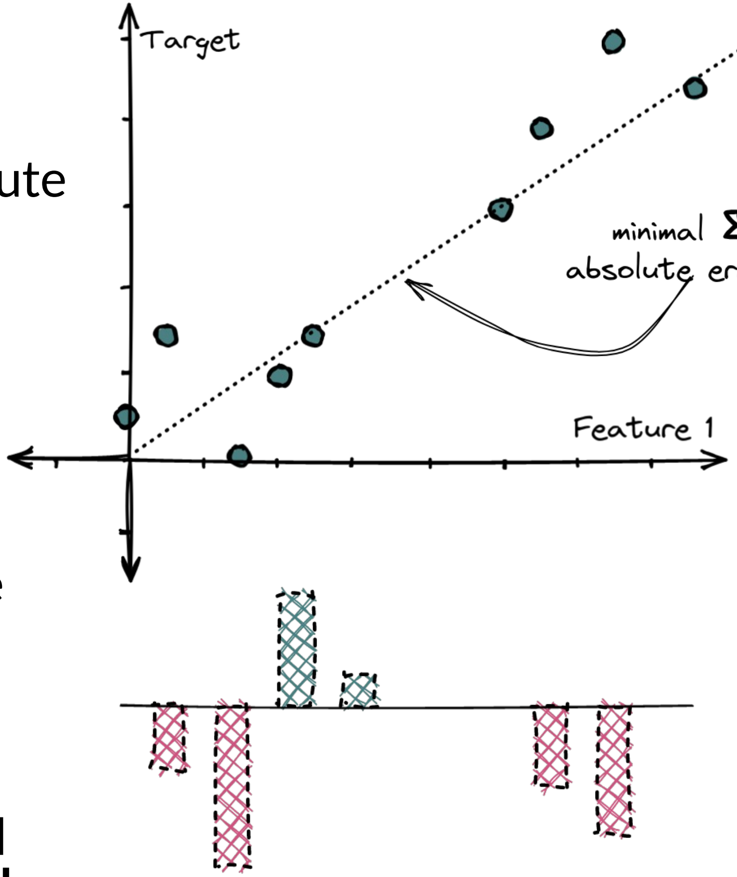


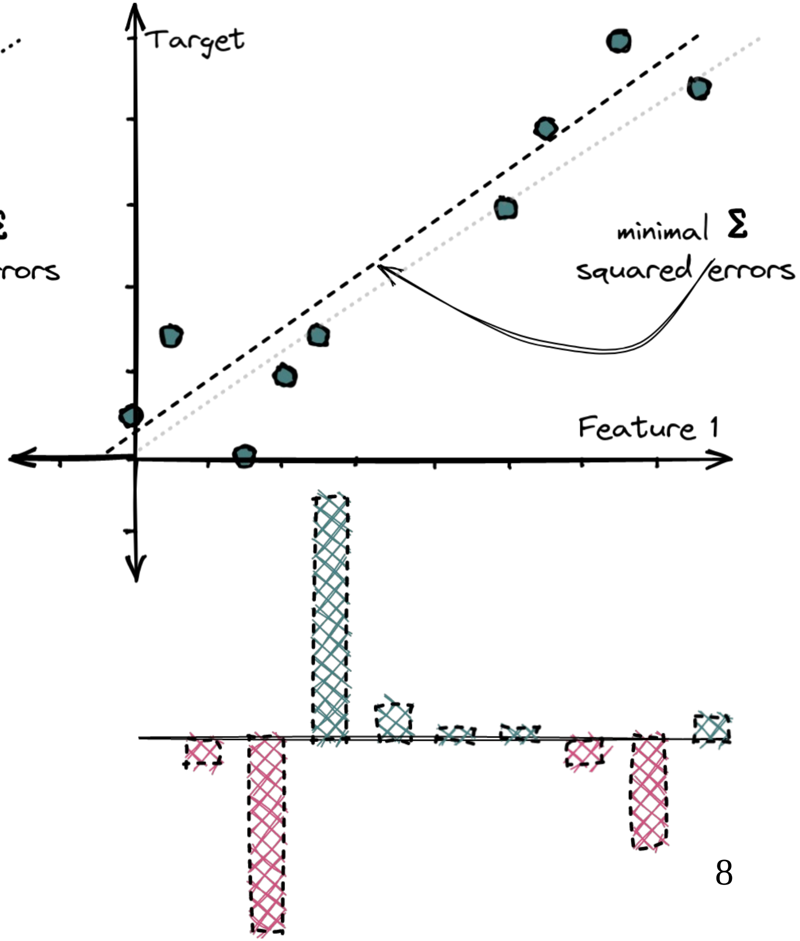
OLS. Step5 : Comparing methods

As one may notice, the distribution of positive/negative errors for the absolute errors is worse compared to the squared errors.

Absolute errors method tends to fit the closest data points, whereas the squared errors method focuses on fitting all points.

This is why **squared errors are preferred**.





OLS. Step6 : Find best fit

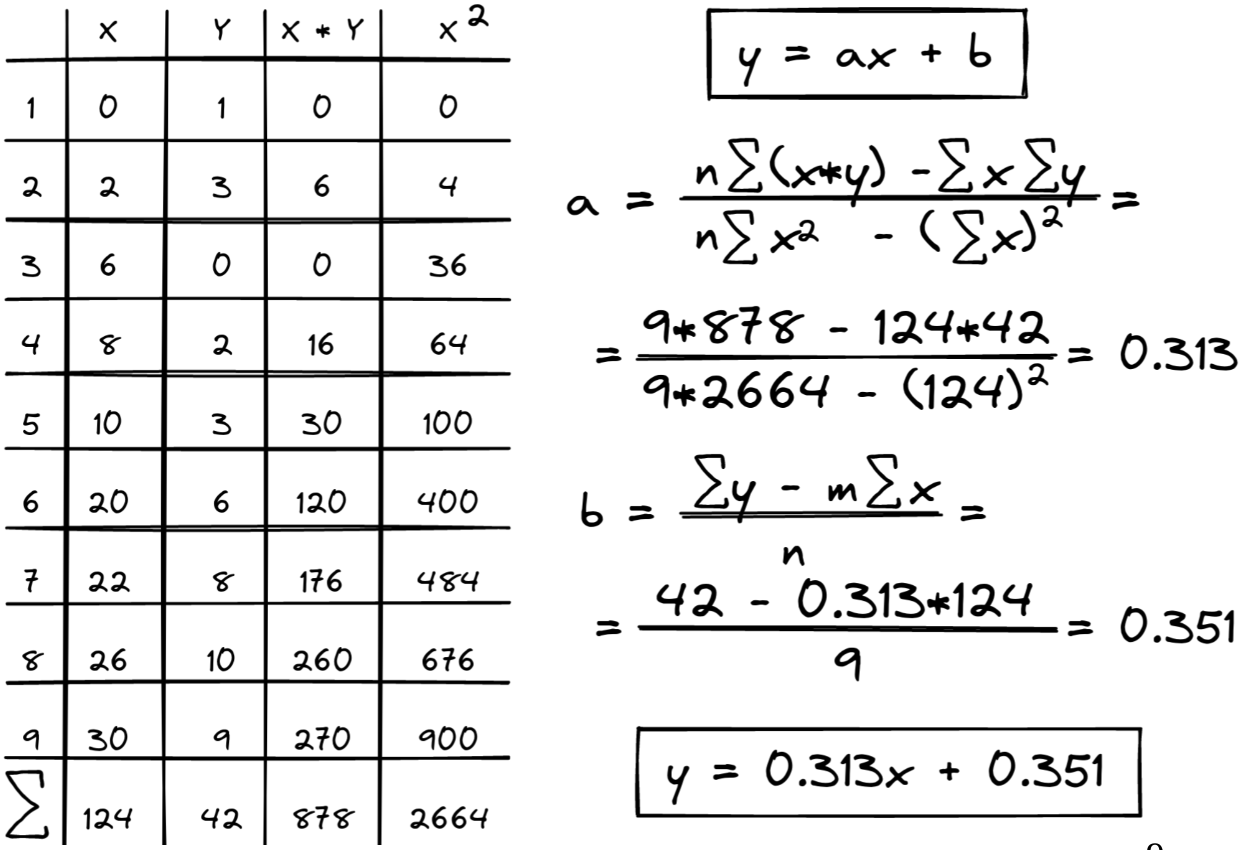
Now that we've decided we're going with squared errors, it's time to find the parameters of the best fitting line.

One way to do that is to draw lines and calculate their fit. In some time, there is a chance we'll find the best fit.

Another way is to use math →

The best fitting line for our data is:

y = 0.313x + 0.351



	X	Y	X * Y	X^2
1	0	1	0	0
2	2	3	6	4
3	6	0	0	36
4	8	2	16	64
5	10	3	30	100
6	20	6	120	400
7	22	8	176	484
8	26	10	260	676
9	30	9	270	900
$\sum$	124	42	878	2664

$$\boxed{y = ax + b}$$

$$a = \frac{n\sum(x*y) - \sum x \sum y}{n\sum x^2 - (\sum x)^2} =$$

$$= \frac{9*878 - 124*42}{9*2664 - (124)^2} = 0.313$$

$$b = \frac{\sum y - m\sum x}{n} =$$

$$= \frac{42 - 0.313*124}{9} = 0.351$$

$$\boxed{y = 0.313x + 0.351}$$