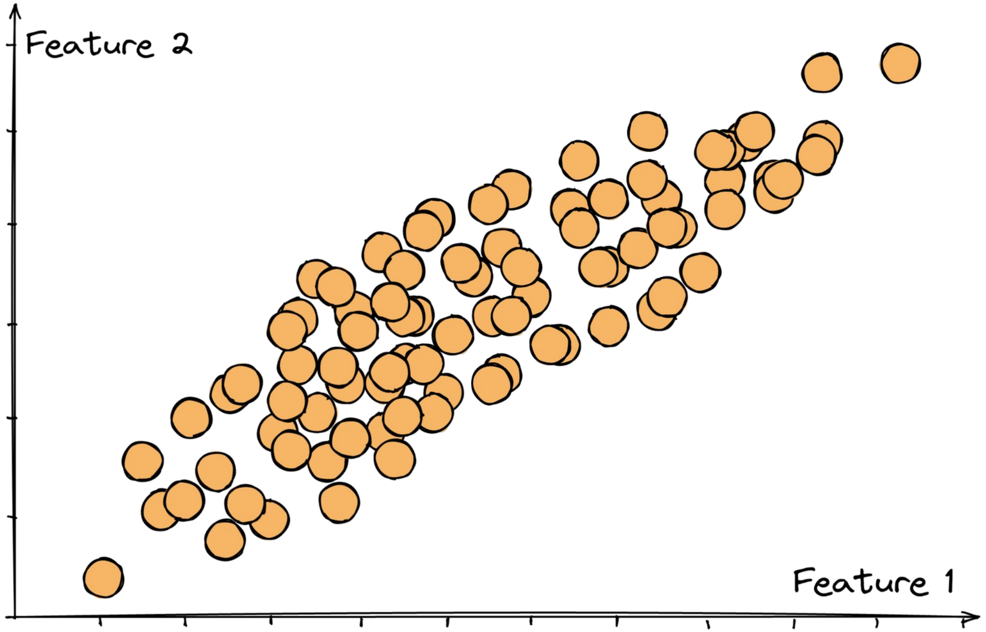
Linear Regression



Week 8

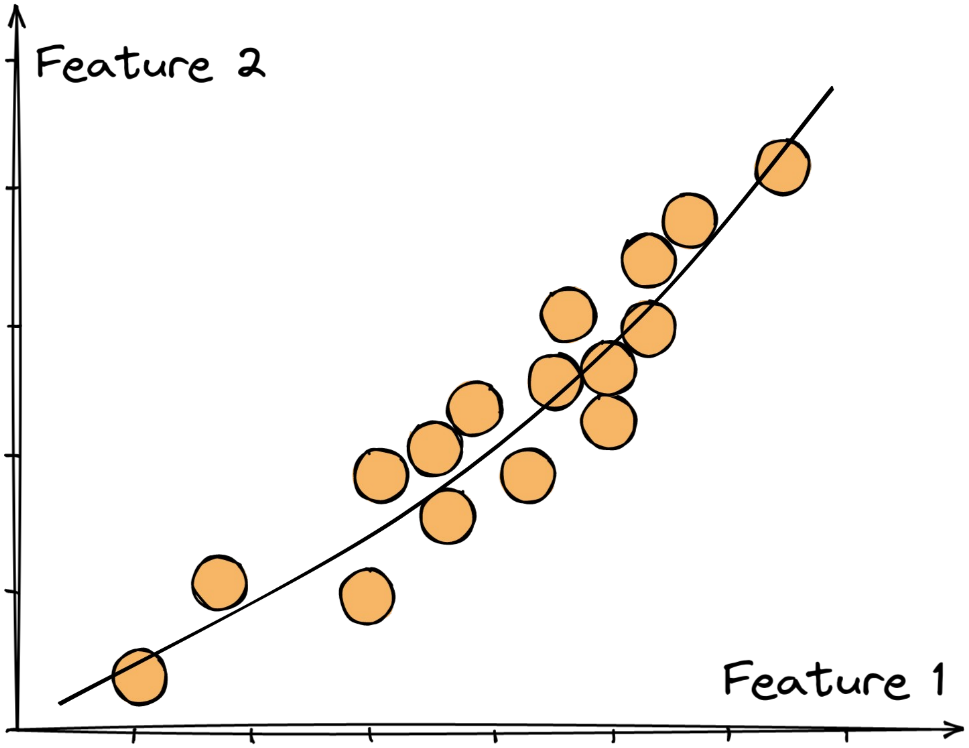
Middlesex University Dubai;
CST4050; Instructor: Ivan Reznikov

Plan

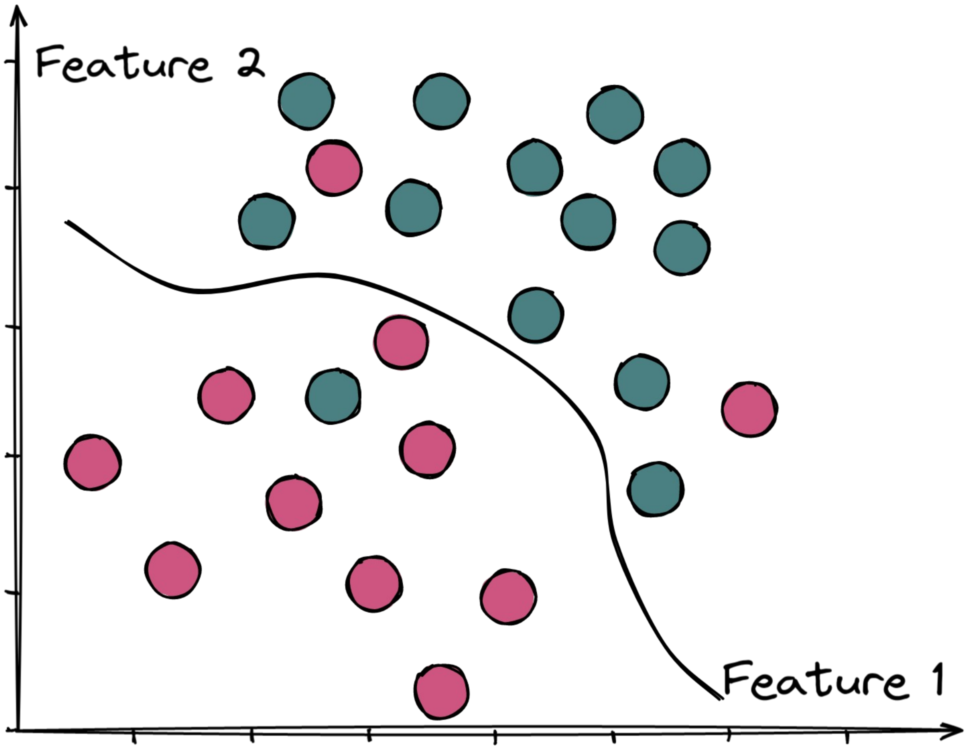
- Main forecasting tasks of Machine Learning
- Supervised/unsupervised learning
- Noise and patterns
- Ordinary least squares
- Evaluation metrics for regression
- Bias-Variance trade-off

Main forecasting tasks of ML

Regression

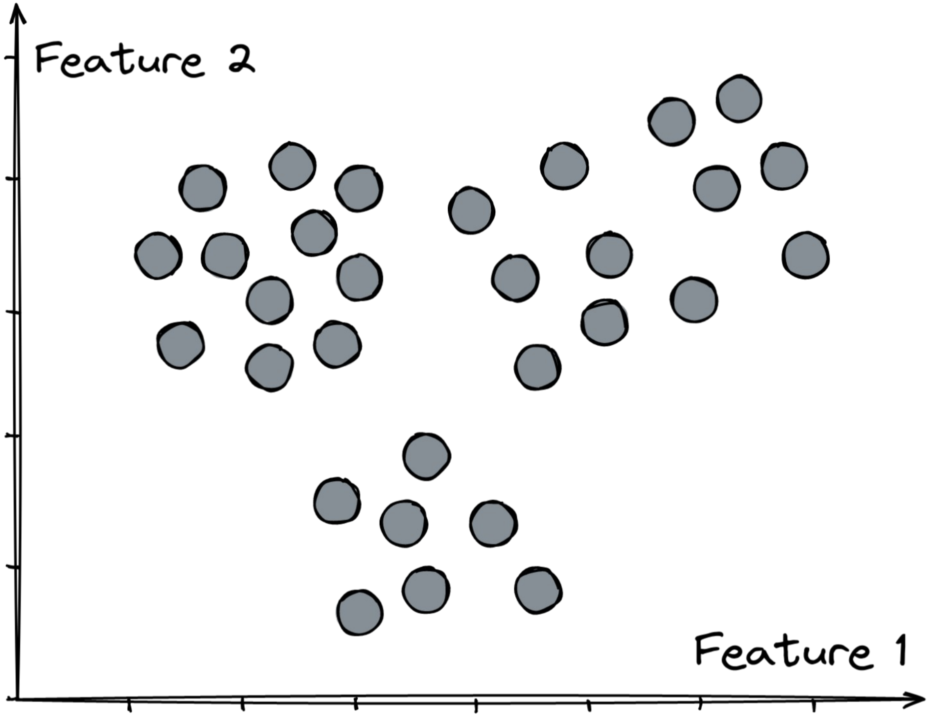


Classification

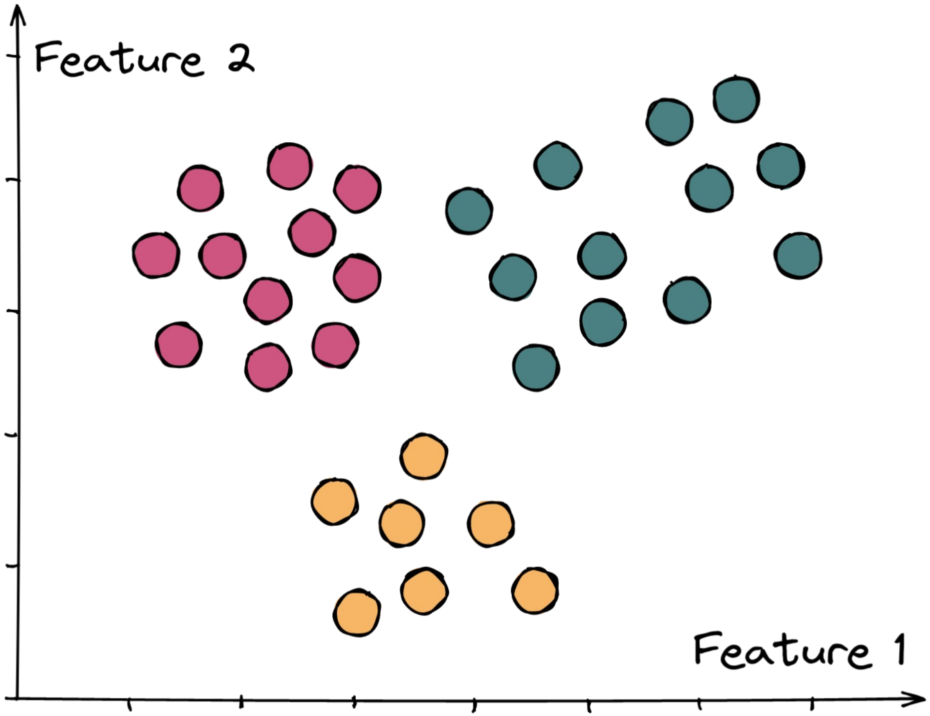


Unsupervised/Supervised

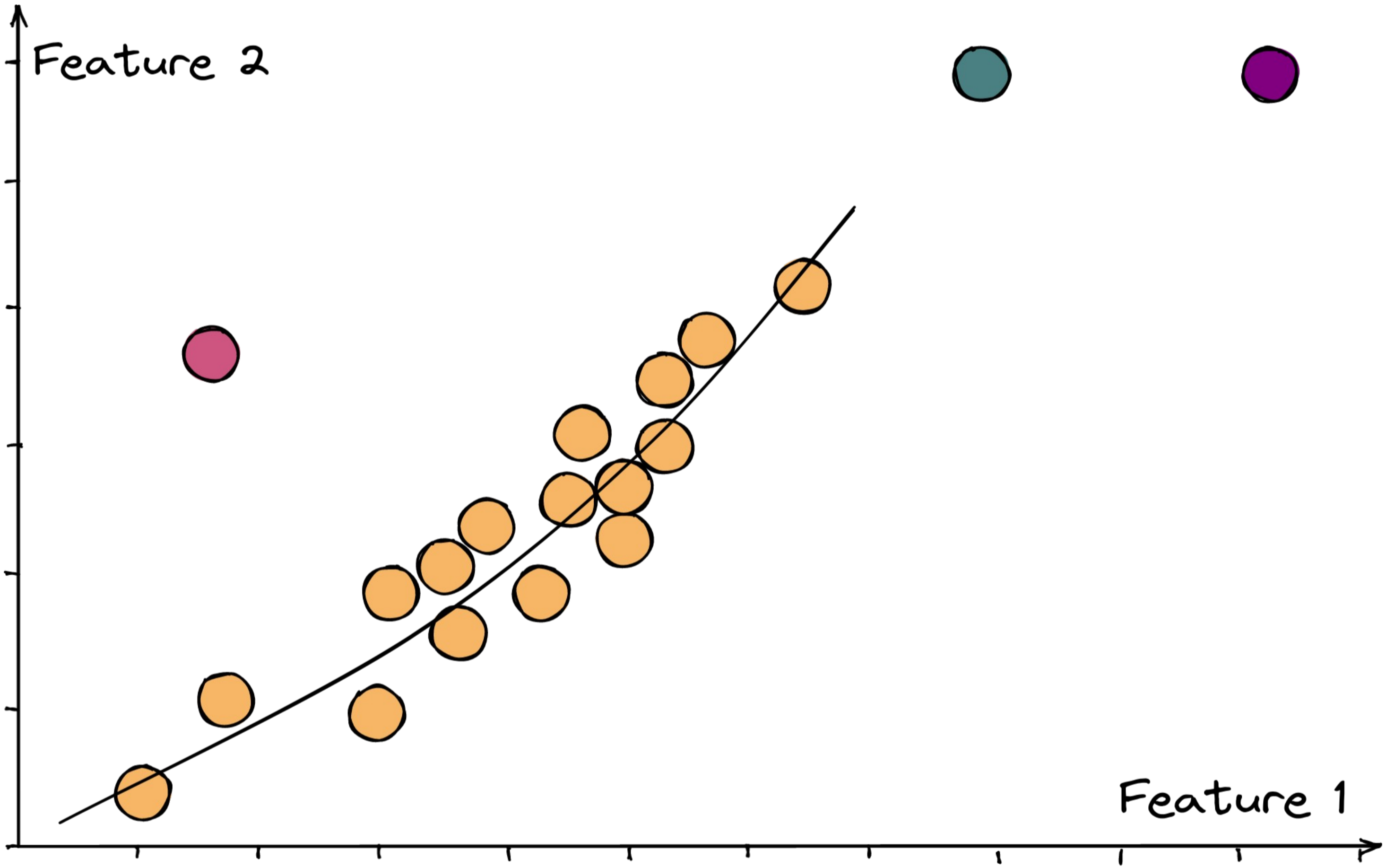
Unlabeled Data



Labeled Data



Patterns and Noise



Ordinary least squares

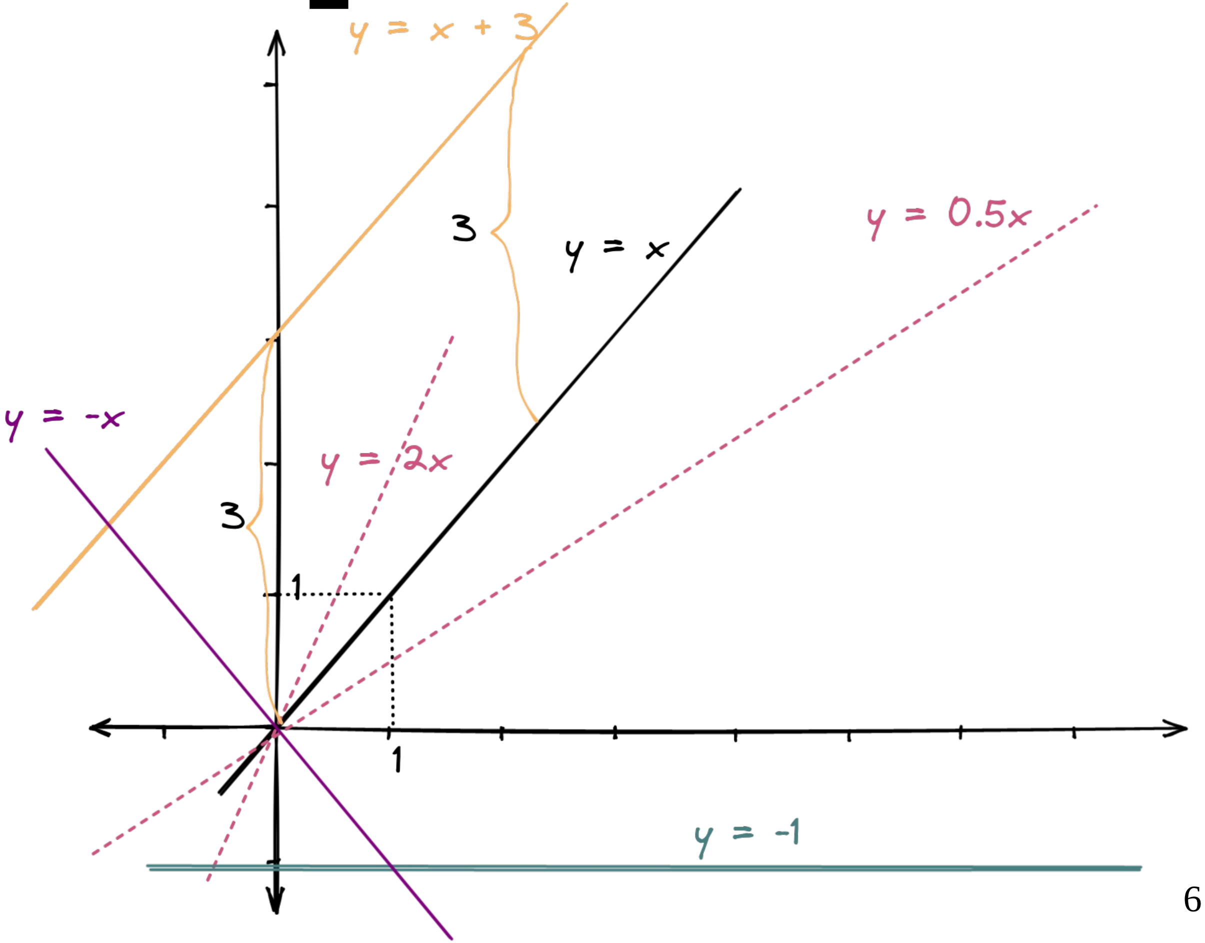
Any line can be described by the following equation:

y = ax + b

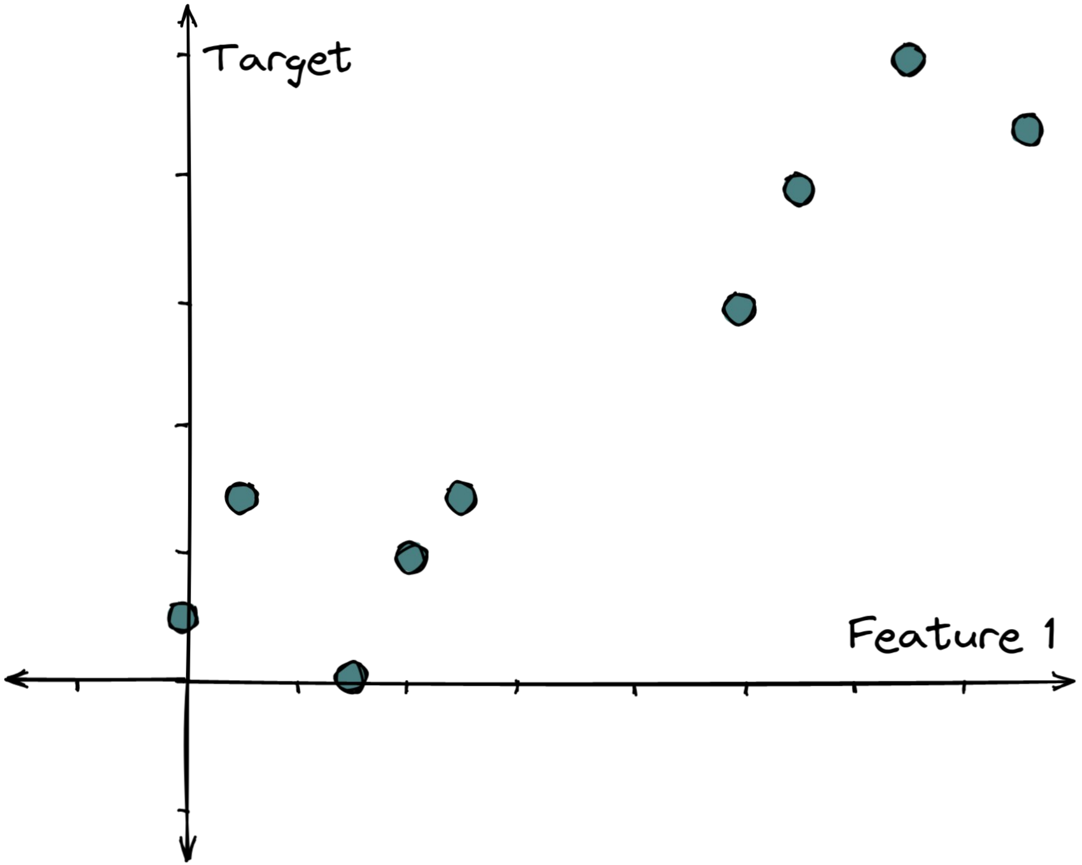
a – slope
b – intercept

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

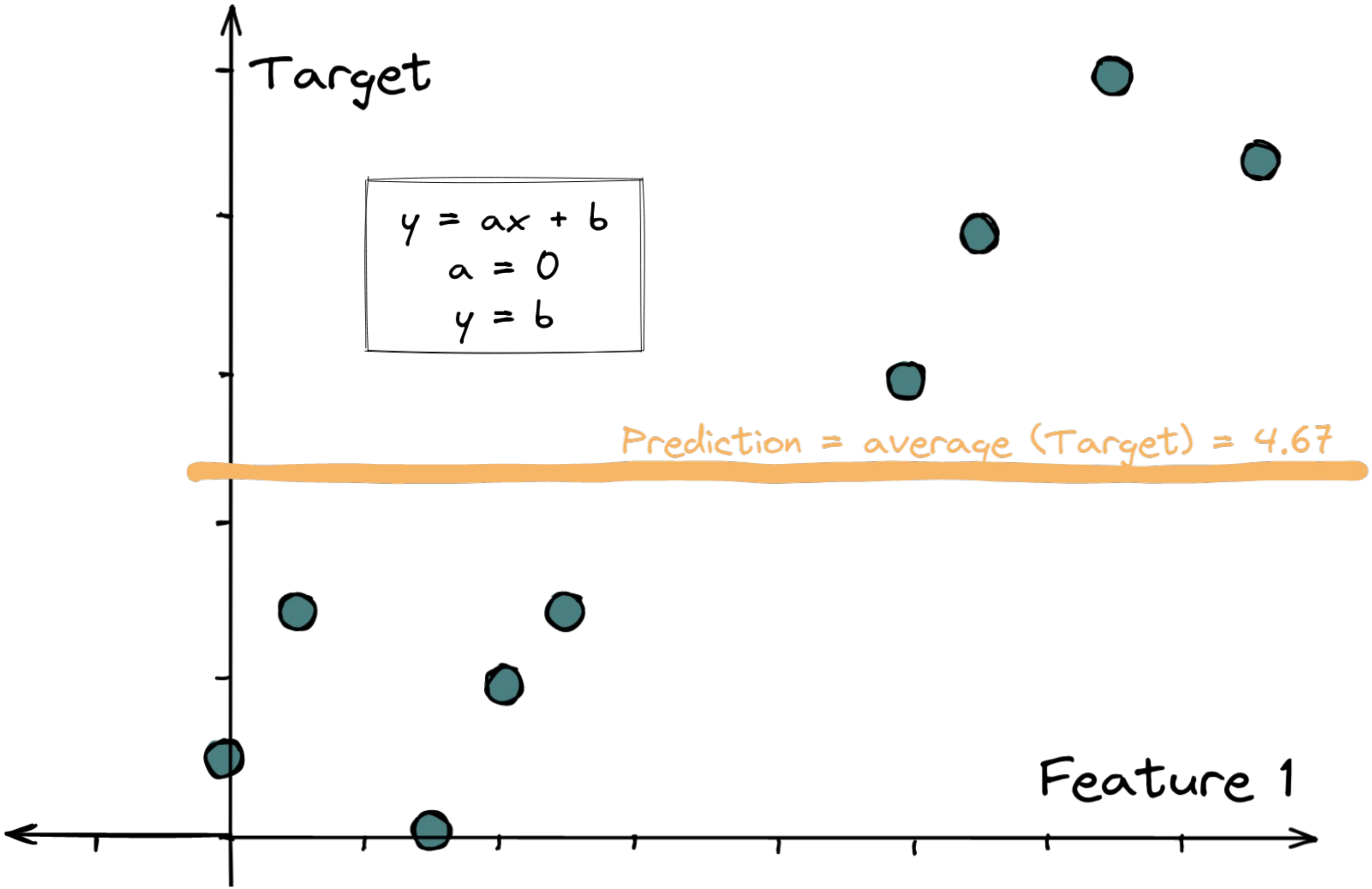


OLS. Step1 : Plot Data



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

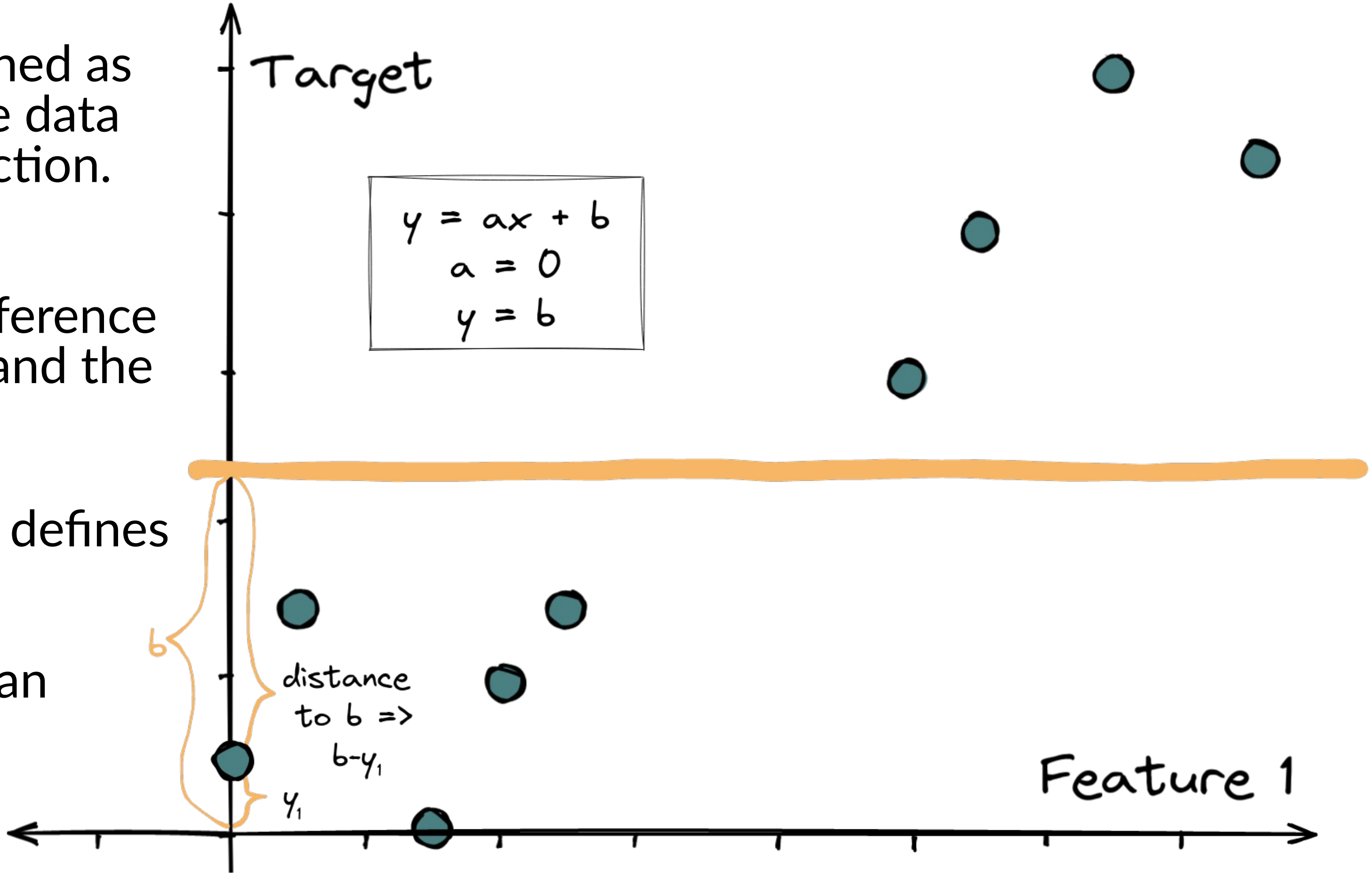
Errors can be defined as distances from the data point to the prediction.

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



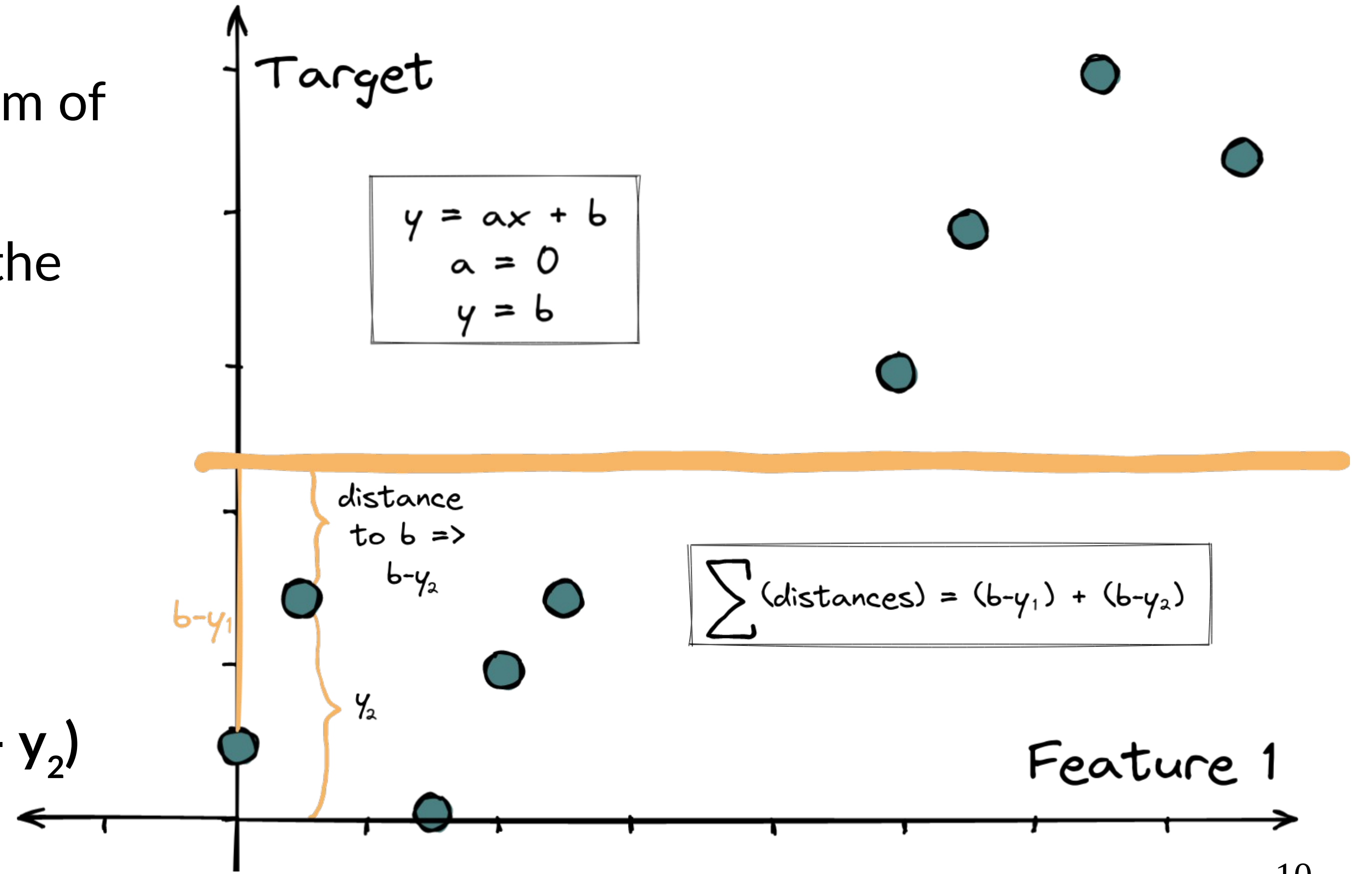
OLS. Step3b : Find errors

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

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

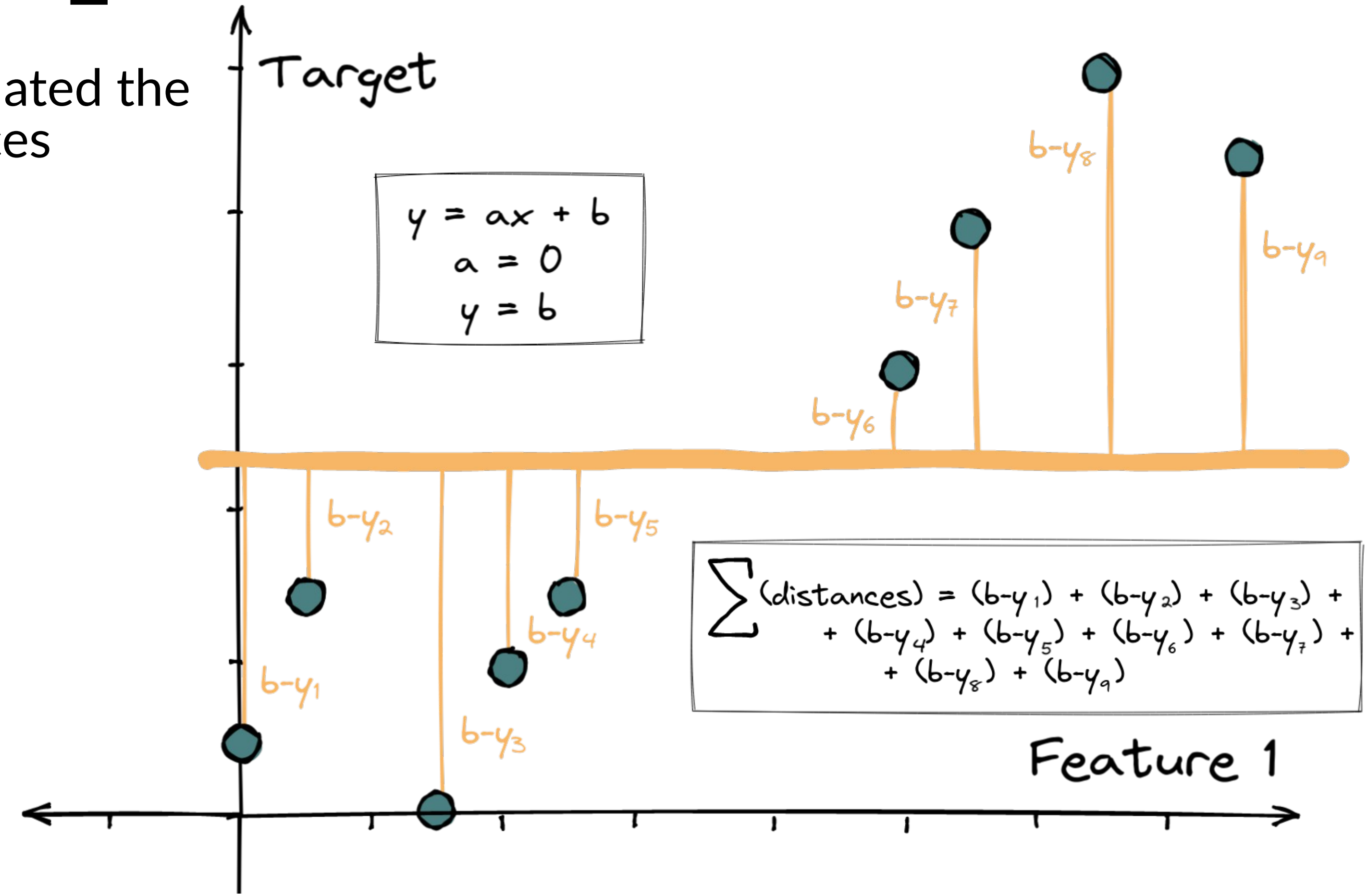
For data point 1: **b – y_1**
For data point 2: **b – y_2**

Current sum: (b – y_1) + (b – y_2)



OLS. Step3c : Find errors

Now we’ve calculated the sum of all distances

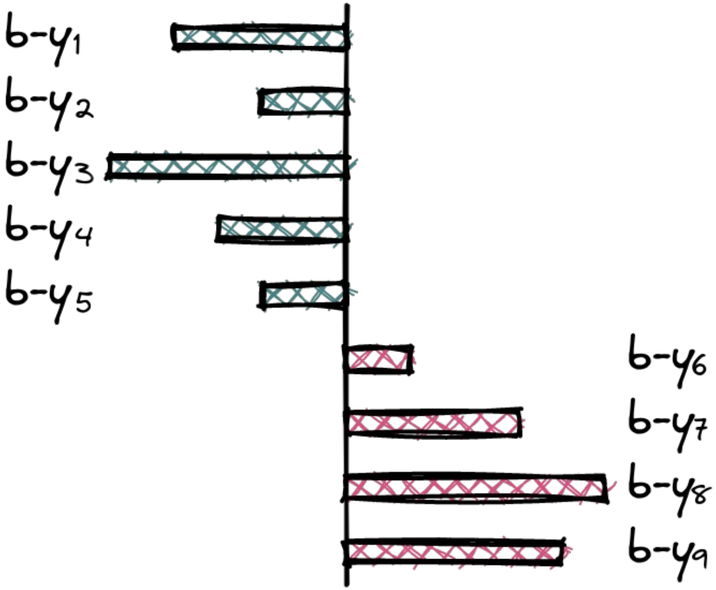


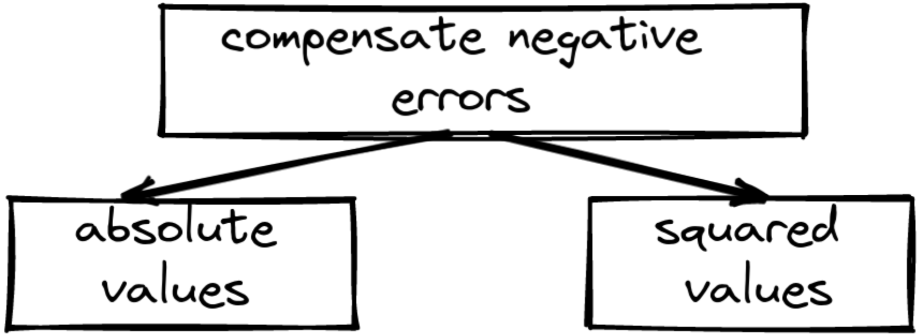
OLS. Step4 : Analyze errors

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.

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



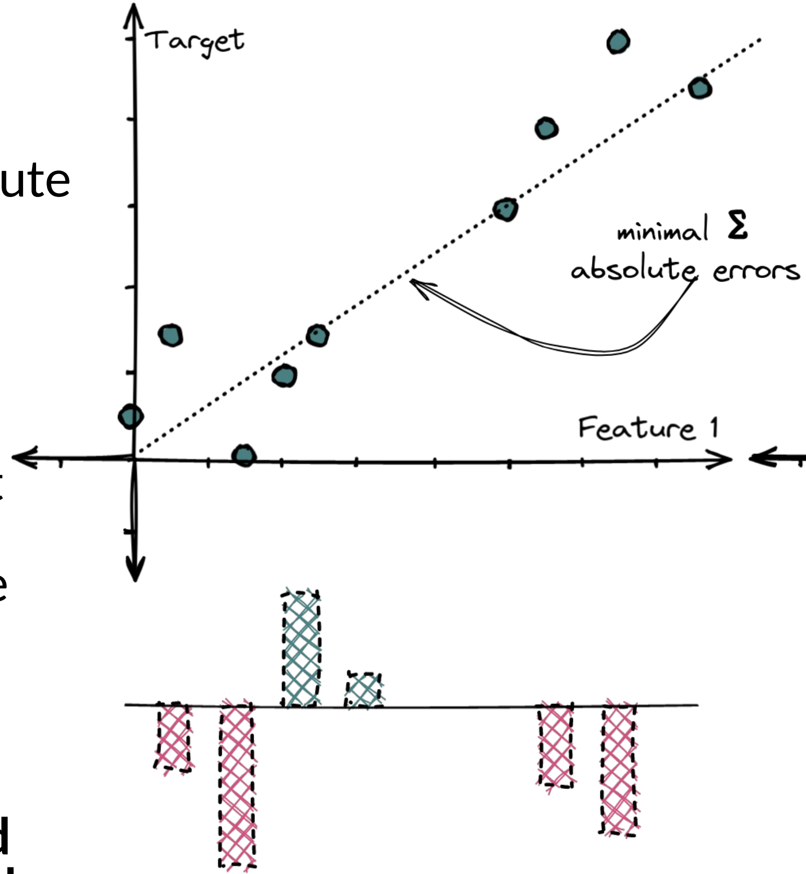


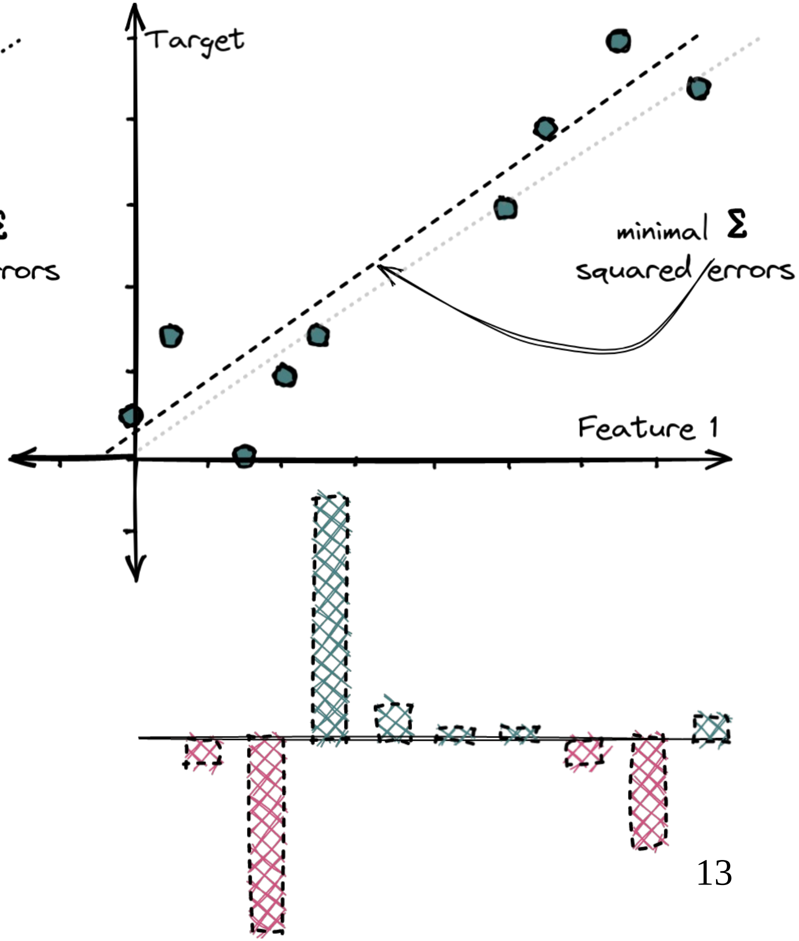
OLS. Step5 : Comparing methods

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

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

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

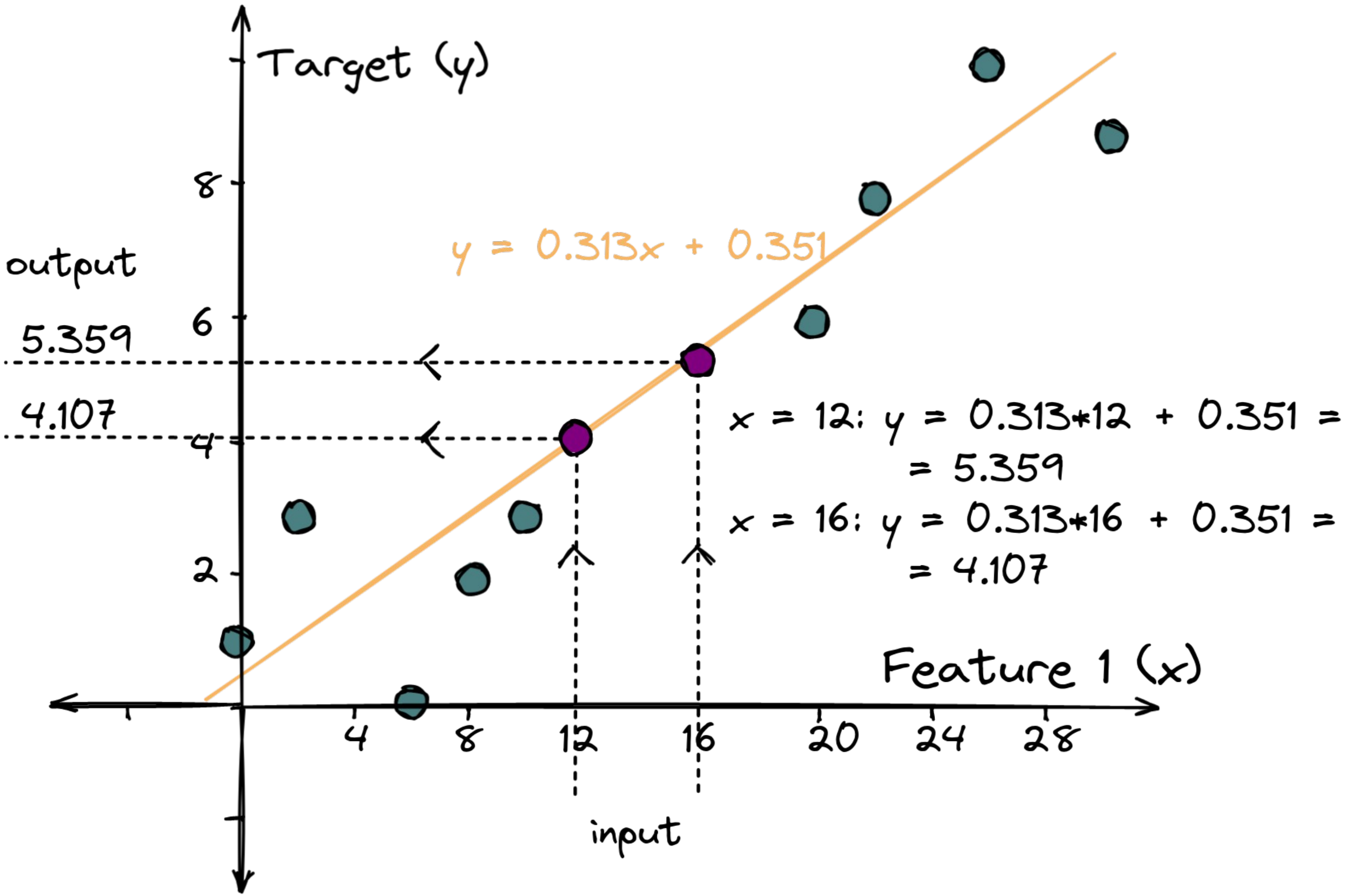
$$= \frac{9*878 - 124*42}{9*2664 - (124)^2} = 0.313$$

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

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

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

OLS. Step7 : Predict



Types of Errors

Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

The most common metric for regression tasks. MSE is calculated by the sum of square of prediction error which is real output minus predicted output, all divide by the number of data points. An absolute number on how much your predicted results deviate from the actual number.

Cannot be interpreted from one single result, but is great for comparison. MSE penalize large errors

Types of Errors

Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE}$$

Mean Square Error (RMSE) is the square root of MSE. It is used more commonly than MSE because firstly sometimes MSE value can be too big to compare easily. Secondly, MSE is calculated by the square of error, and thus square root brings it back to the same level of prediction error and makes it easier for interpretation.

Types of Errors

Mean Absolute Error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Compare to MSE or RMSE, MAE is a more direct representation of sum of error terms. MSE gives larger penalization to big prediction error by square it while MAE treats all errors the same.

MAE does not penalize large errors.

Types of Errors

R^2:

SS – sum of squares

$$R^2 = 1 - \frac{SS\ (Regr)}{SS\ (Total)} = 1 - \frac{\sum_i (y_i - \hat{y})^2}{\sum_i (y_i - \bar{y})^2}$$

R^2 measures how much variability in dependent variable can be explained by the model. R Square is calculated by the sum of squared of prediction error divided by the total sum, which replaces the calculated prediction with mean. R Square value is between 0 to 1, and a greater value indicates a better fit between forecast and actual value. R^2 is an excellent measure to determine how well the model fits the dependent variables. However, it does not take into consideration of overfitting problem.

Types of Errors

$R^2_{adjusted}$:

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - m - 1}$$

n – number of observations

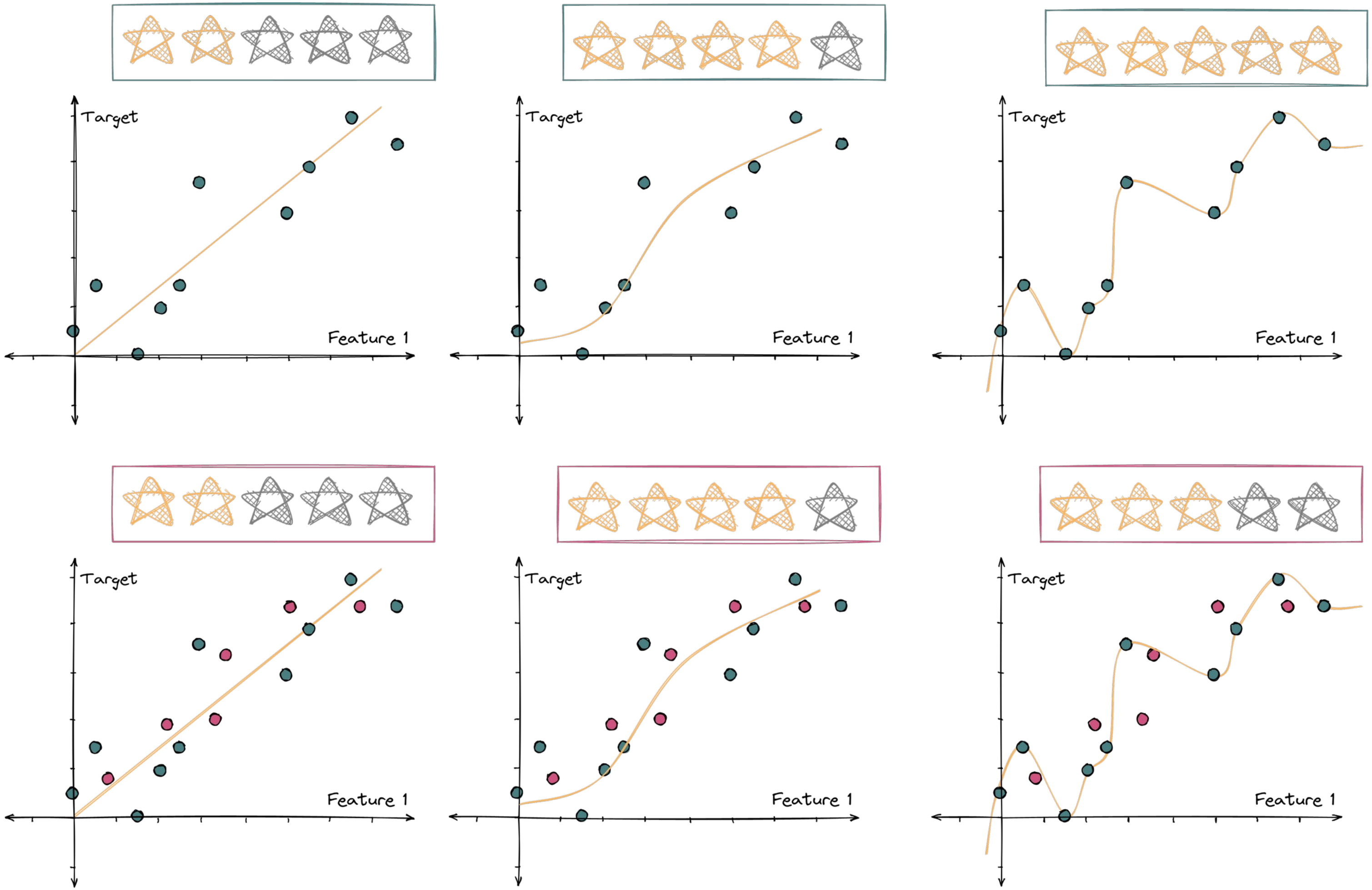
m – number of independent

Adjusted R^2 is the only metric here that considers the overfitting problem.

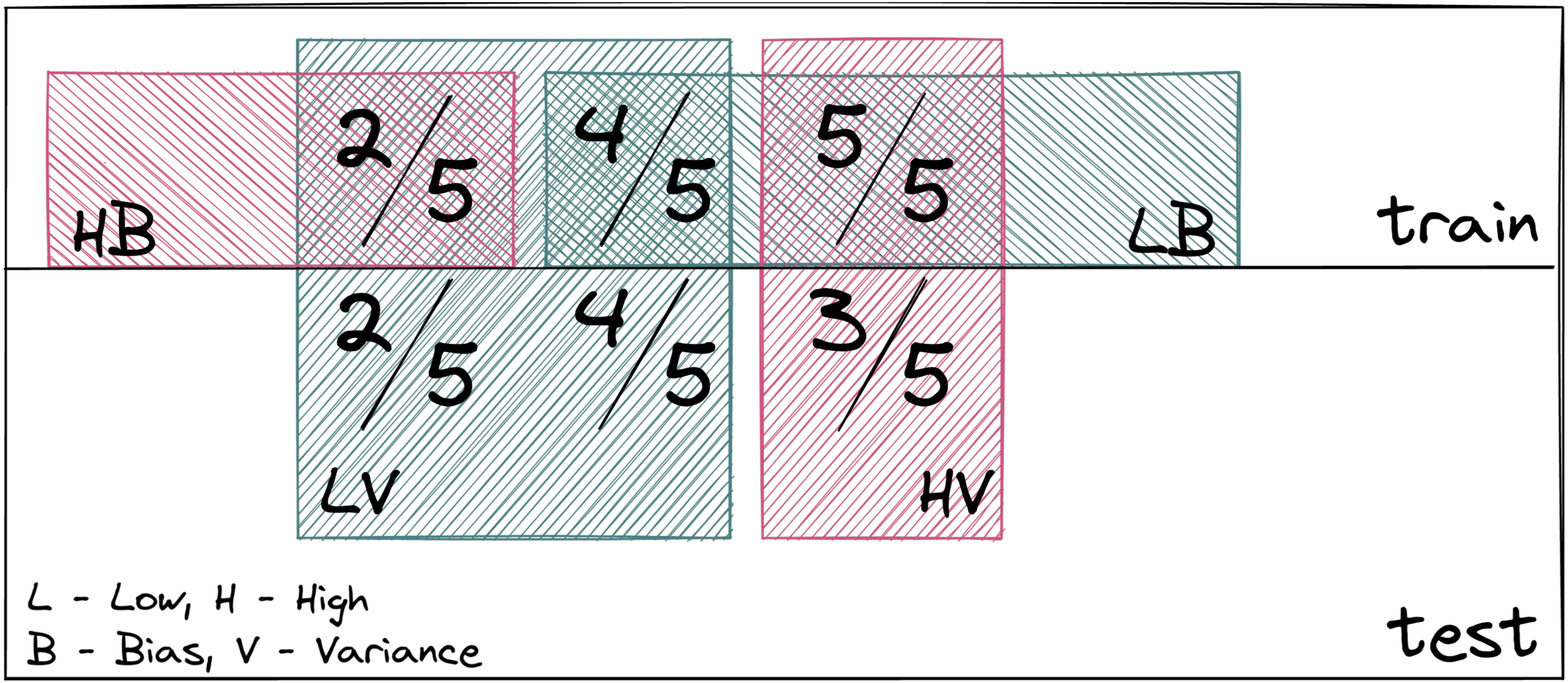
If a new added variable is uncorrelated, then R^2 is not increased and penalty is added. Otherwise, R^2 is increased and overcome the m value.

Bias/Variance

Bias/Variance



Bias/Variance



Bias/Variance

Bias is responsible for quality of the model. It is how good you can describe your training data.

Variance is responsible for the reproducibility of your model on the test dataset.

In supervised ML, since the exact mapping function is unknown, real bias/variance terms can't be determined.

Yet the trade-off helps analyze ML algorithms and control their predictive performance.