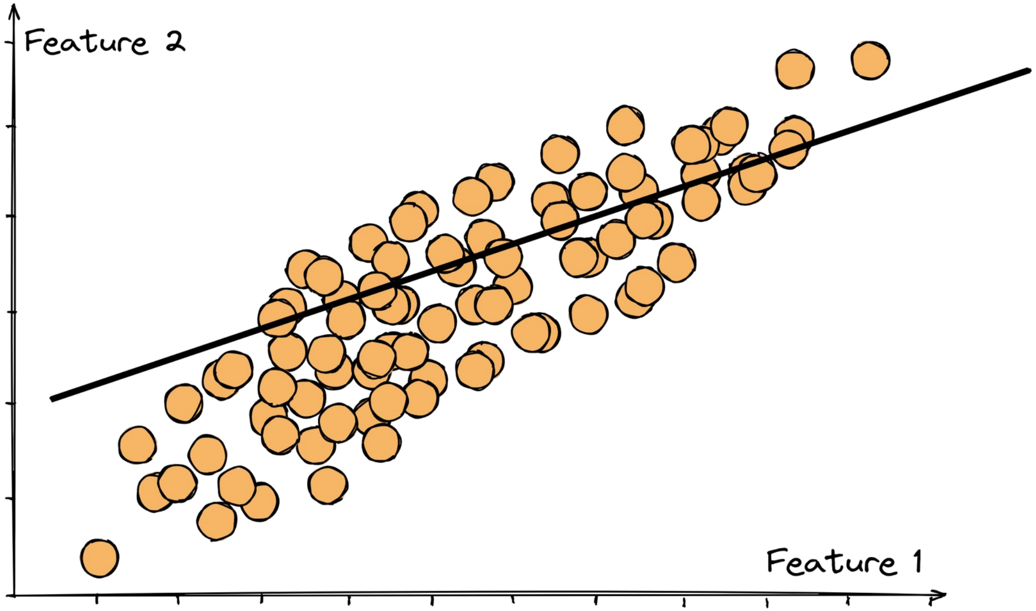
Linear Regression. Part 2



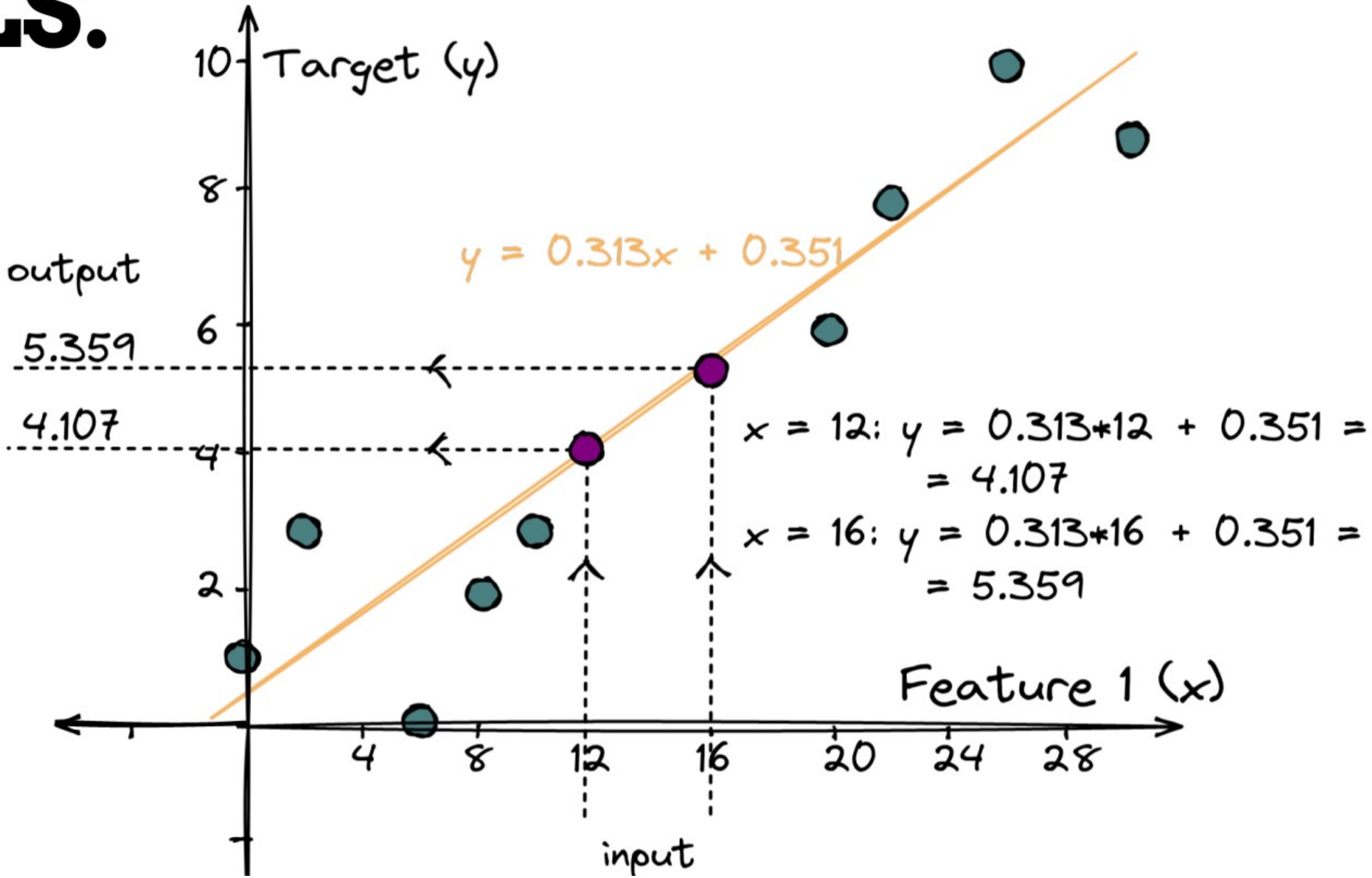
Week 9

Middlesex University Dubai;
CST4050; Instructor: Ivan Reznikov

Plan

- Refreshing Week 8:
  - Ordinary least squares
  - Evaluation metrics for regression
  - Bias-Variance trade-off
- Cross-Validation
- Multiple Linear Regression
- Ridge and Lasso Regularization

OLS.



Target (y)
10
8
6
4
2
output
5.359
4.107
y = 0.313x + 0.351
x = 12: y = 0.313*12 + 0.351 =
= 4.107
x = 16: y = 0.313*16 + 0.351 =
= 5.359
Feature 1 (x)
4
8
12
16
20
24
28
input

Plan

- Refreshing Week 8:
  - Ordinary least squares
  - Evaluation metrics for regression
  - Bias-Variance trade-off
- Cross-Validation
- Multiple Linear Regression
- Ridge and Lasso Regularization

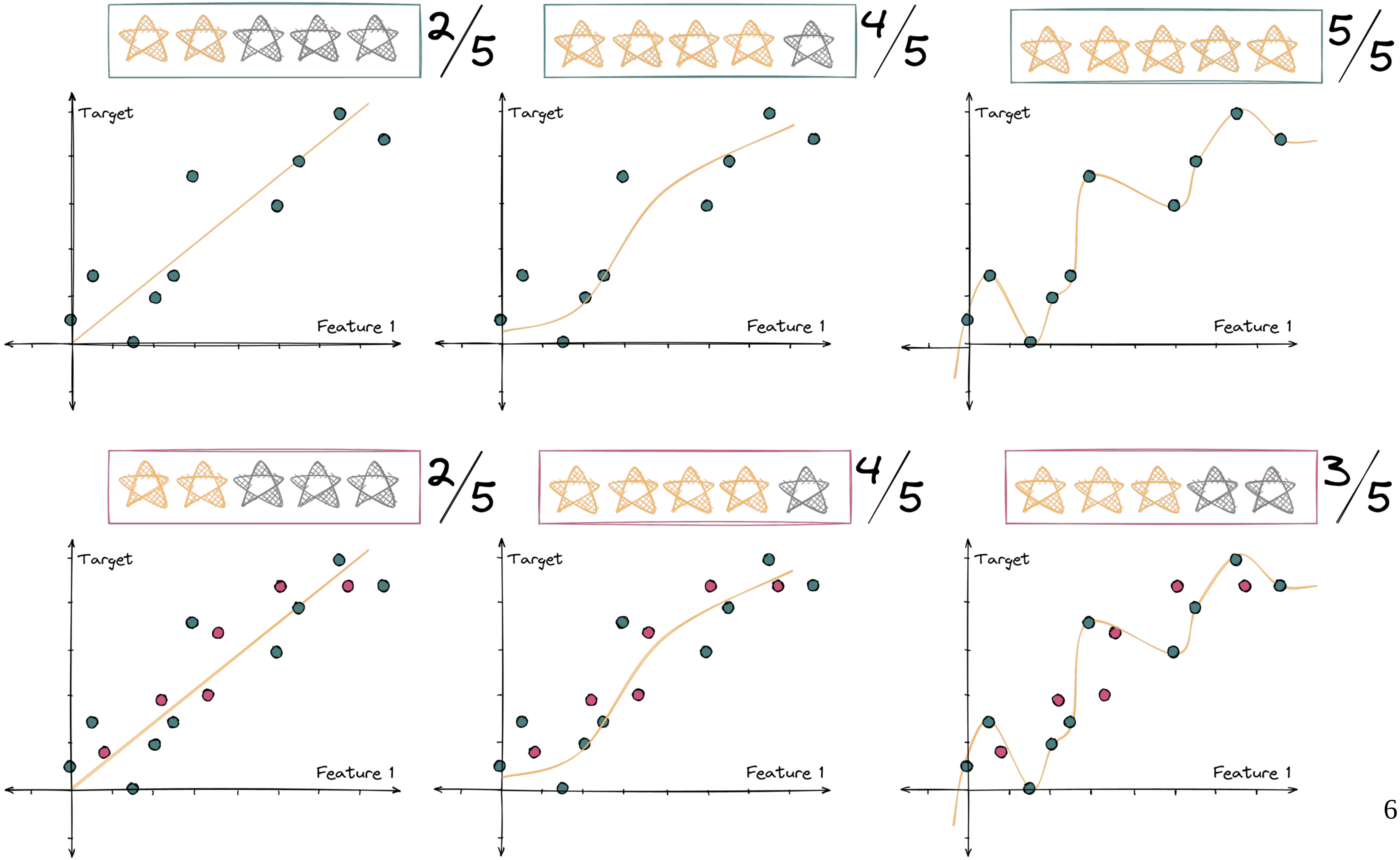
Types of Errors

Mean Squared Error (MSE):

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

Mean Absolute Error (MAE):

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


2/5
Target
Feature 1
4/5
Target
Feature 1
5/5
Target
Feature 1
2/5
Target
Feature 1
4/5
Target
Feature 1
3/5
Target
Feature 1

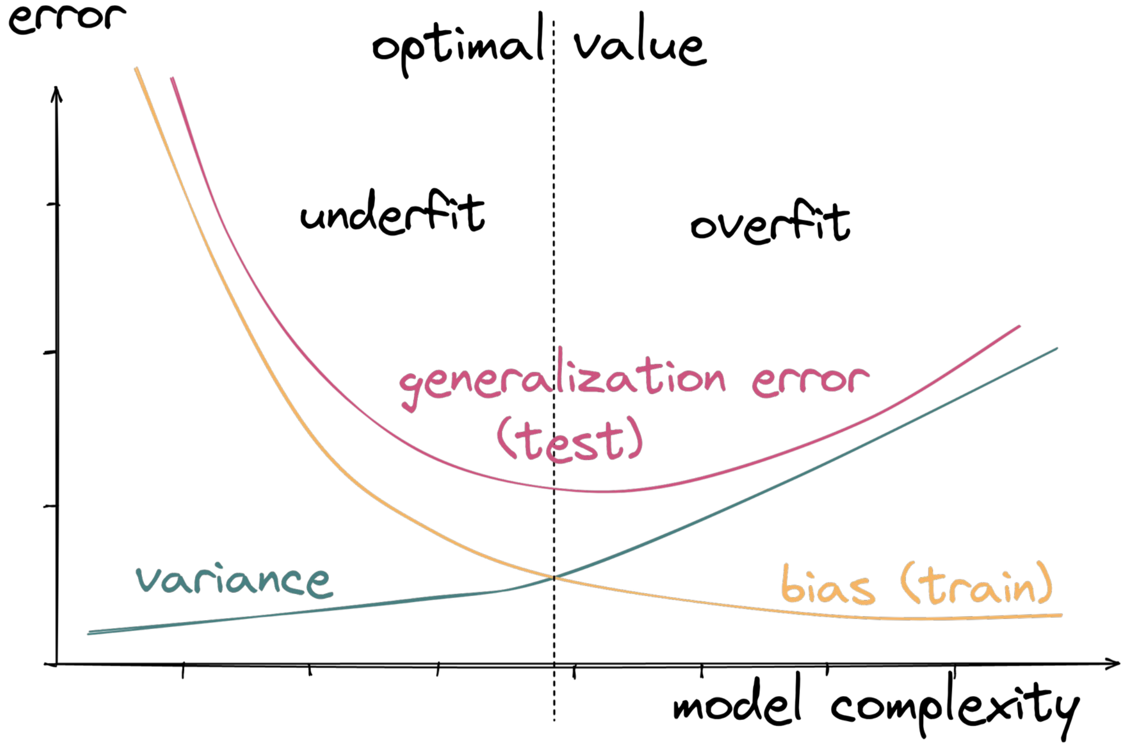
Bias/Variance

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

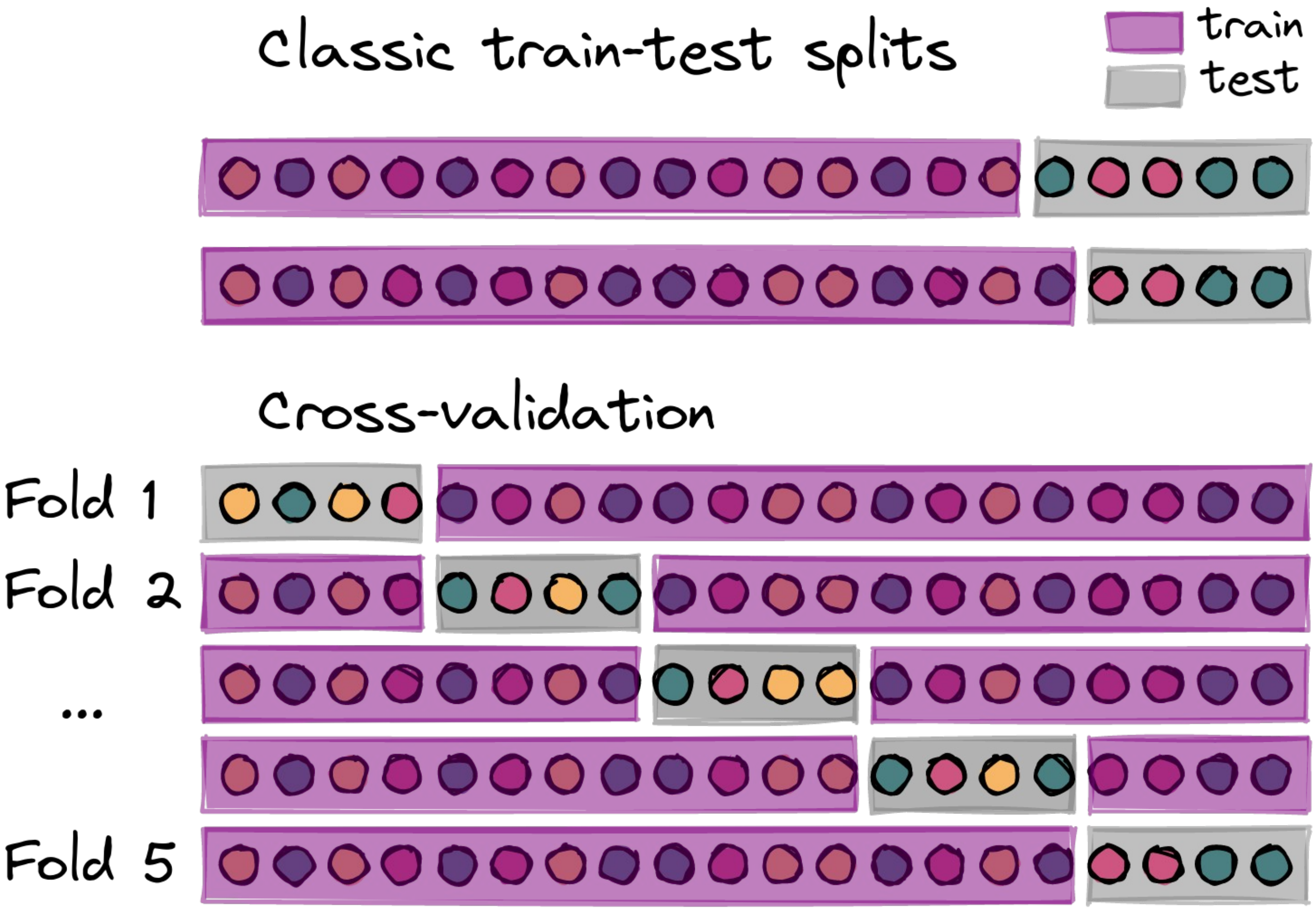
Variance is responsible for the reproducibility of your model on the test dataset.

In supervised ML, real bias/variance terms can't be determined.

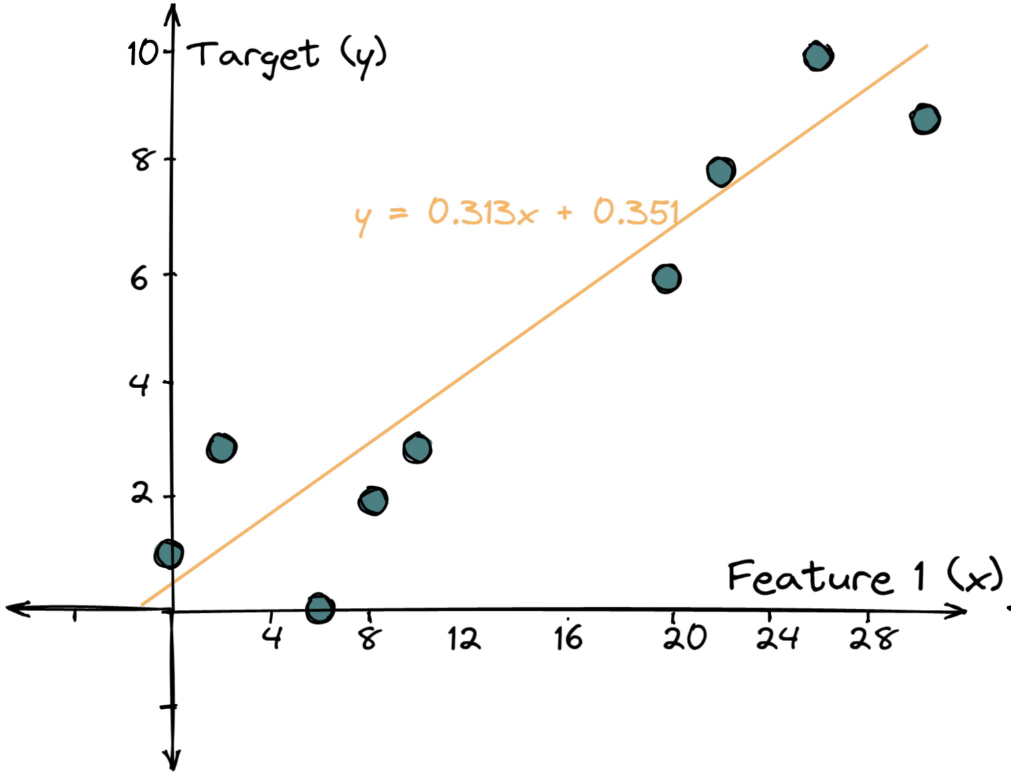
Yet, the trade-off helps analyze ML algorithms and control their predictive performance.

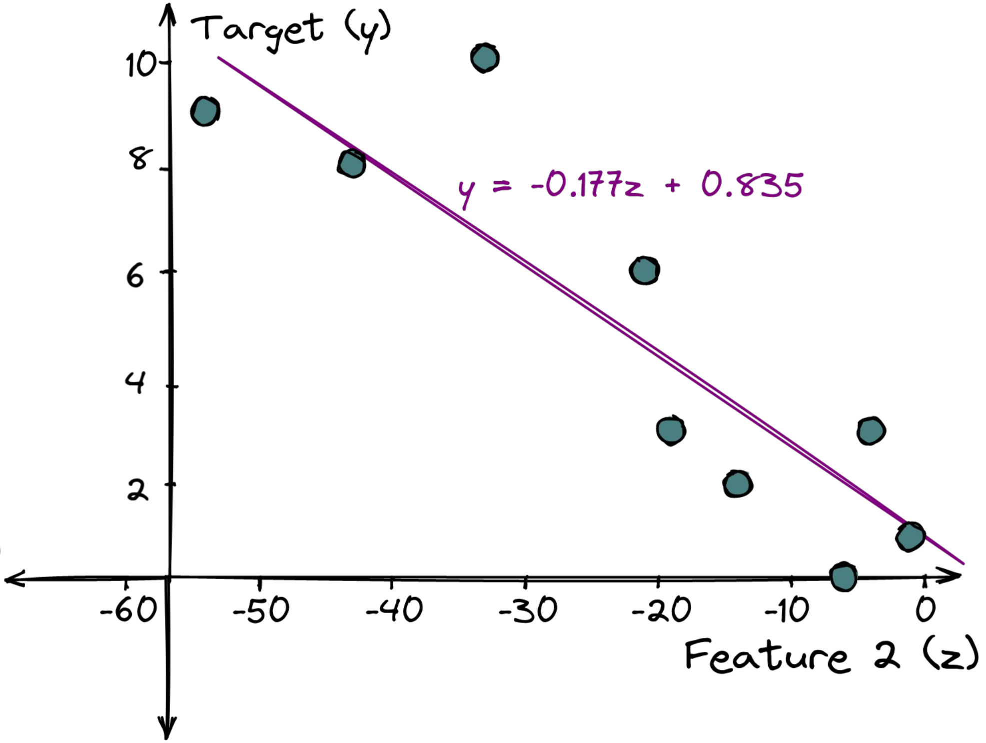


Cross-Validation



Multiple Linear Regression



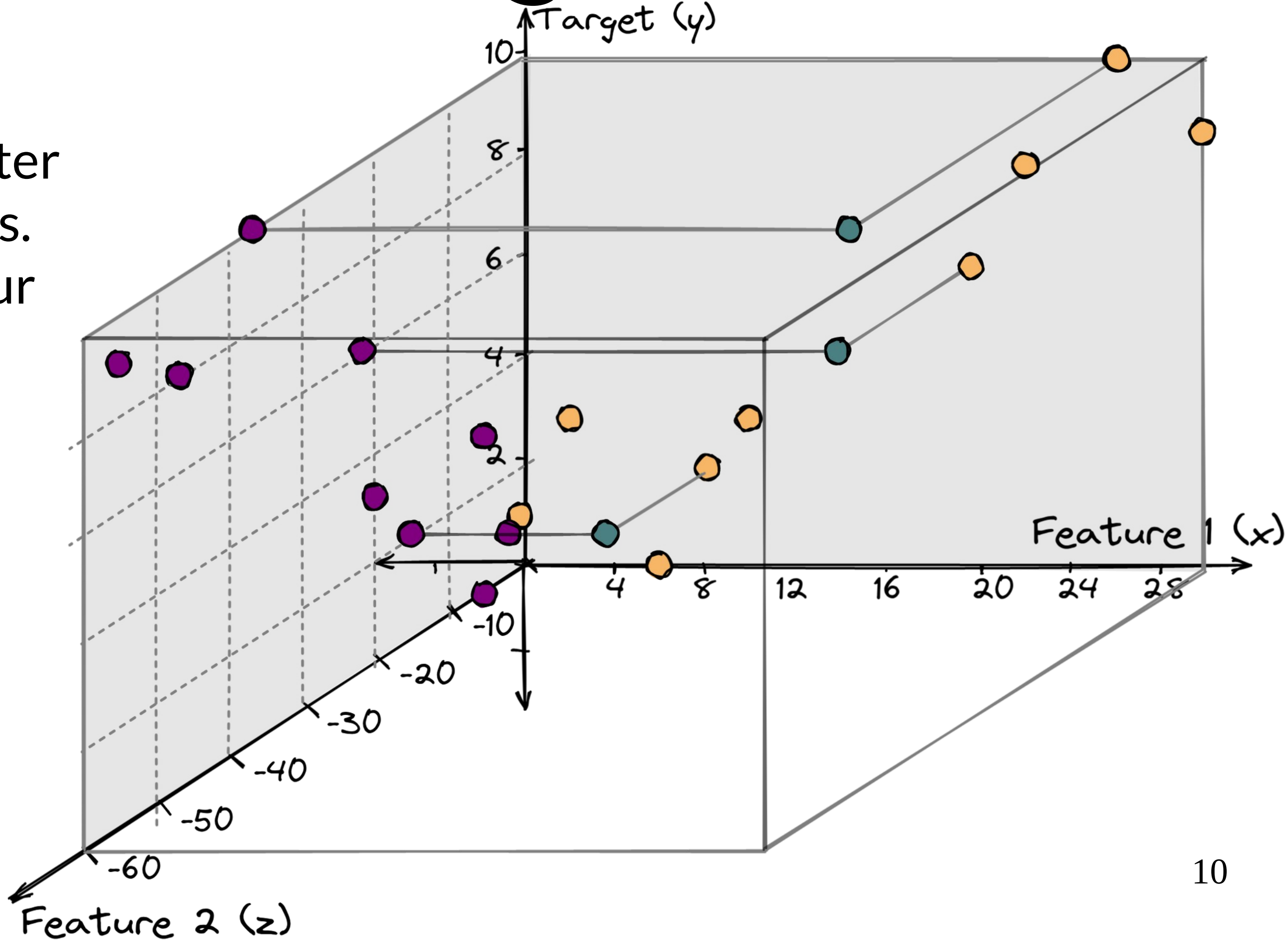


How can we perform several linear regressions at once?

Multiple Linear Regression

We can plot xy and zy scatter plots on appropriate planes. These are projections of our coordinates in 3D space.

Though the points are located in 3D, it is still possible to fit linear regression through them.



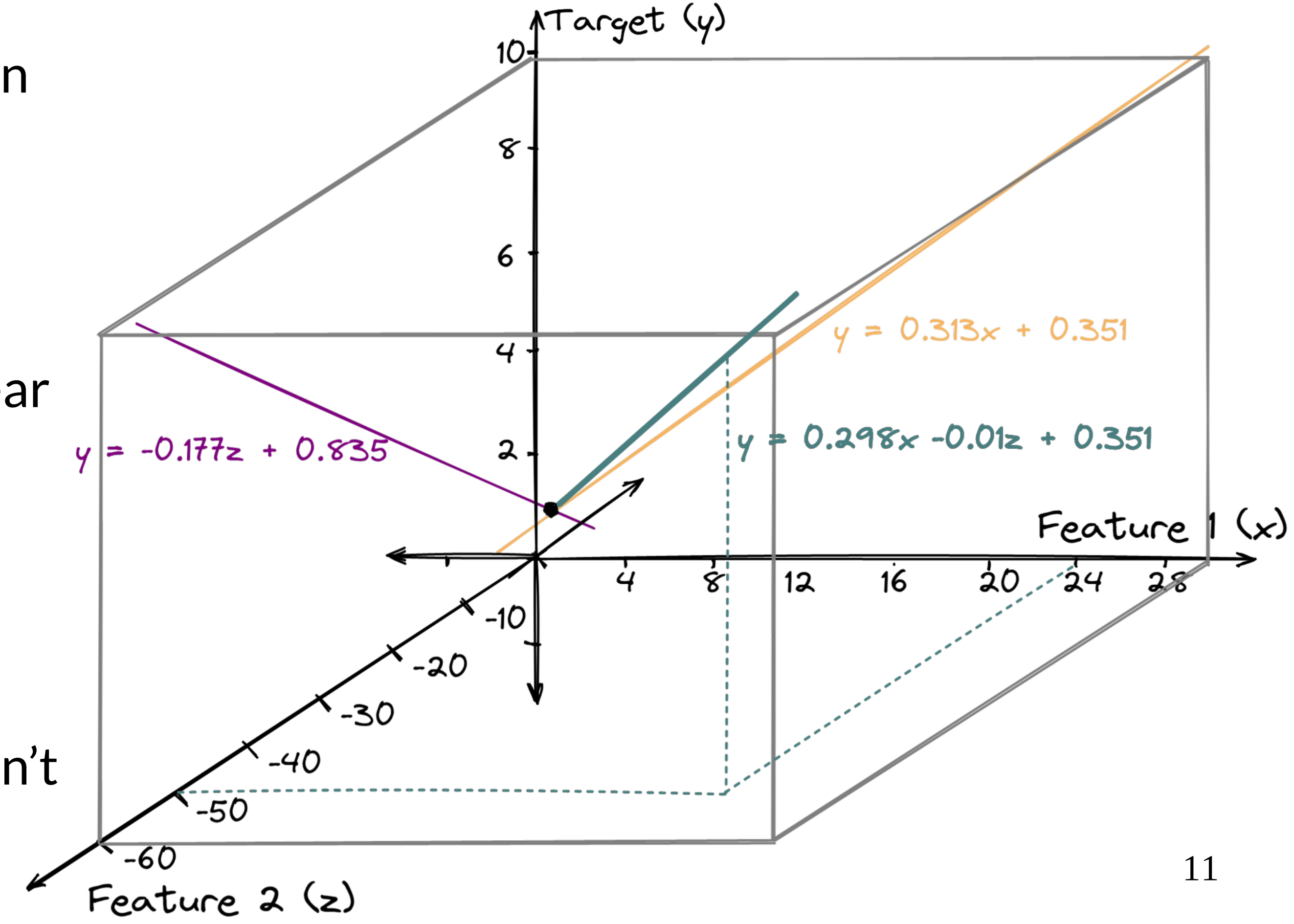
Multiple Linear Regression

In our case, the final equation will be:

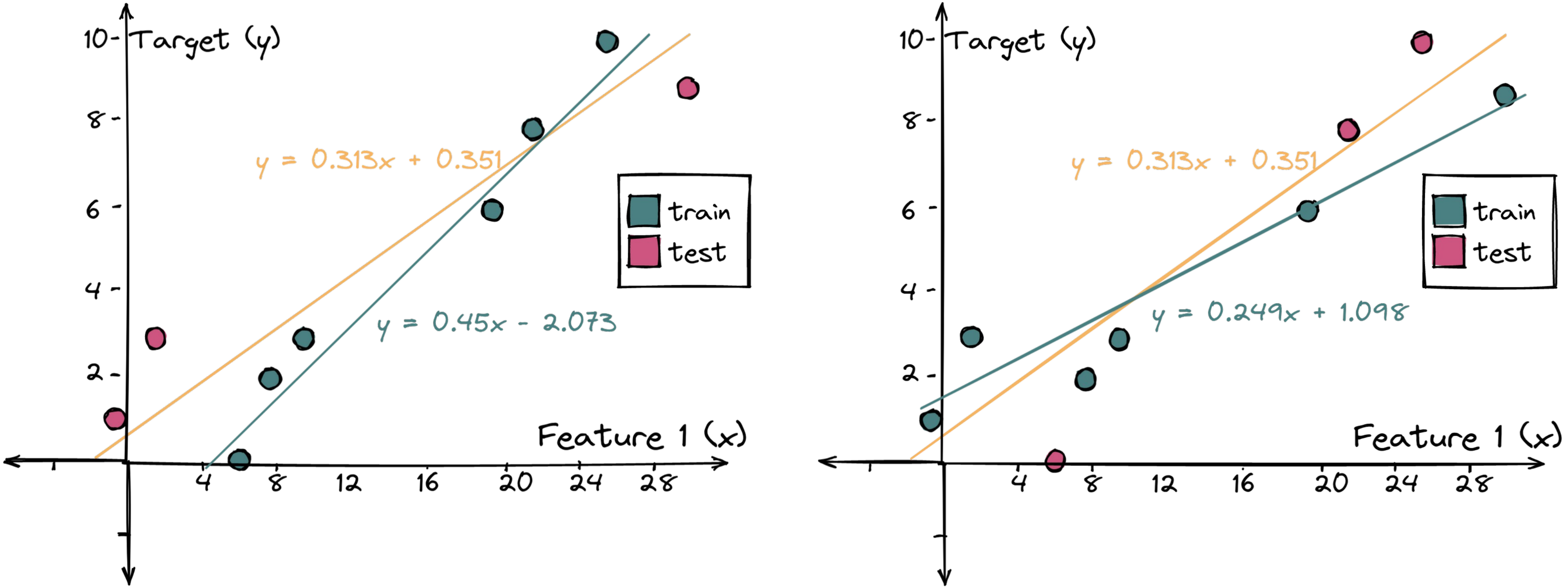
y = 0.298x – 0.01z + 0.351

Notice, that for multiple linear regression we have several slopes for different axes.

Of course, we can visualize only 3 dimensions, but from the math point of view, doesn't matter if there are 10 dimension.

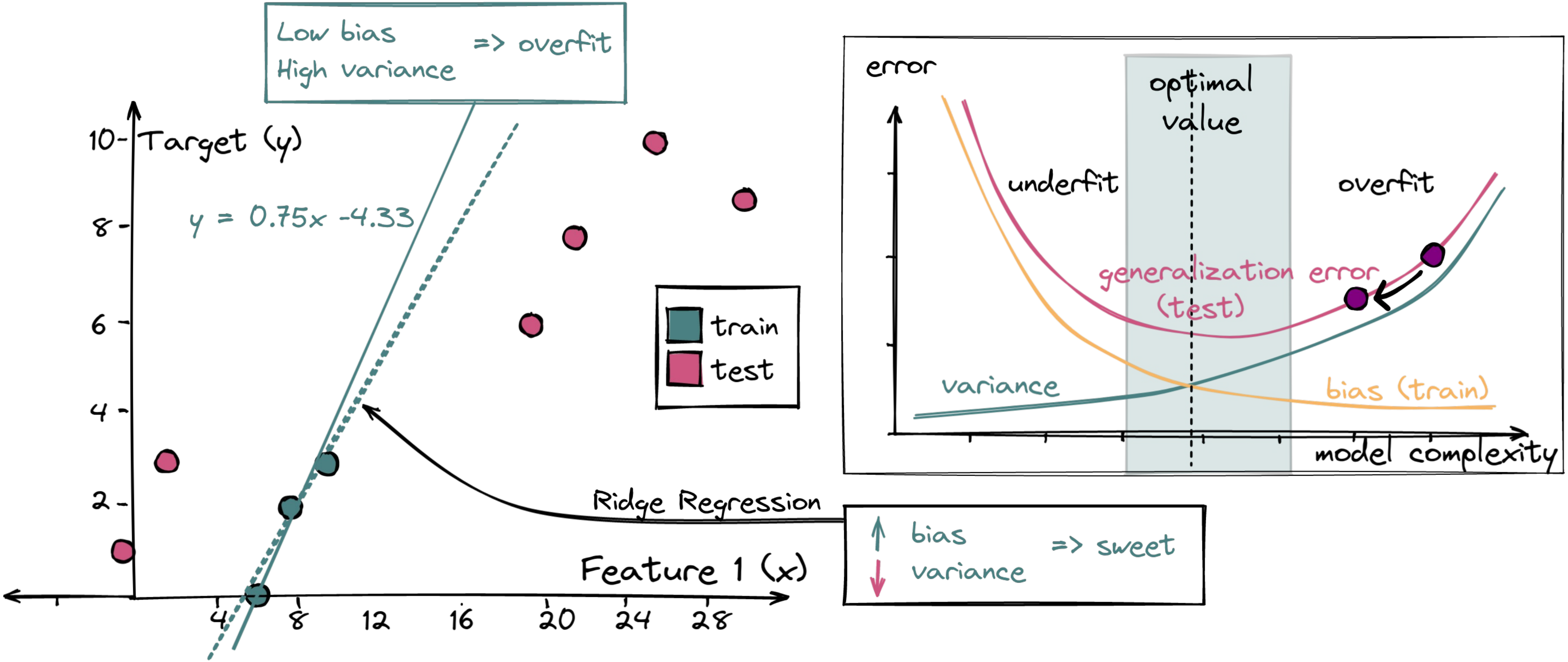


High Variance Example



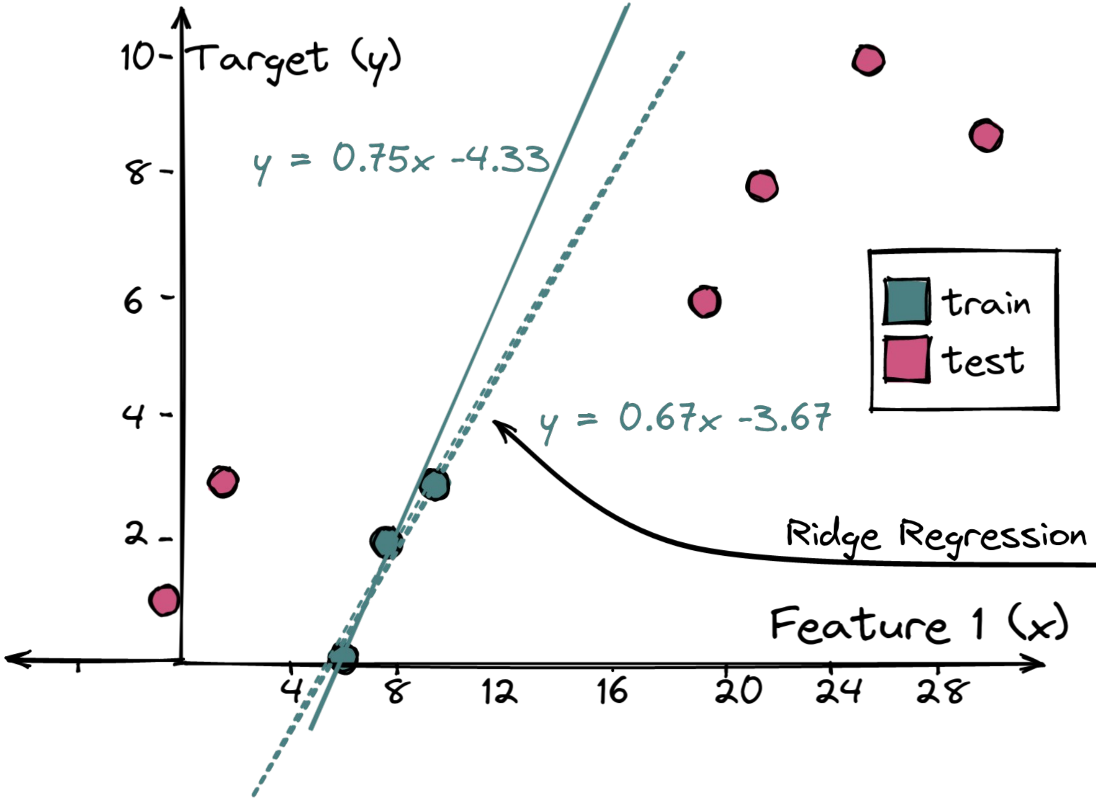
Purpose of both Ridge and Lasso Regularization is to make y less sensible to x

Regularization: Ridge



Ridge (L2) Regularization

minimizing sum of squared residuals



$y = ax + b$

L2 = sum_sq_residuals + $\lambda * a^2$

(assume λ = 1):

$L2(orig) = (-0.167)^2 + 0.33^2 + (-0.167)^2 + 1 * 0.75^2 = \boxed{0.727178}$

$L2(new) = 0.33^2 + (-0.33)^2 + 0^2 + 1 * 0.67^2 = \boxed{0.6667}$

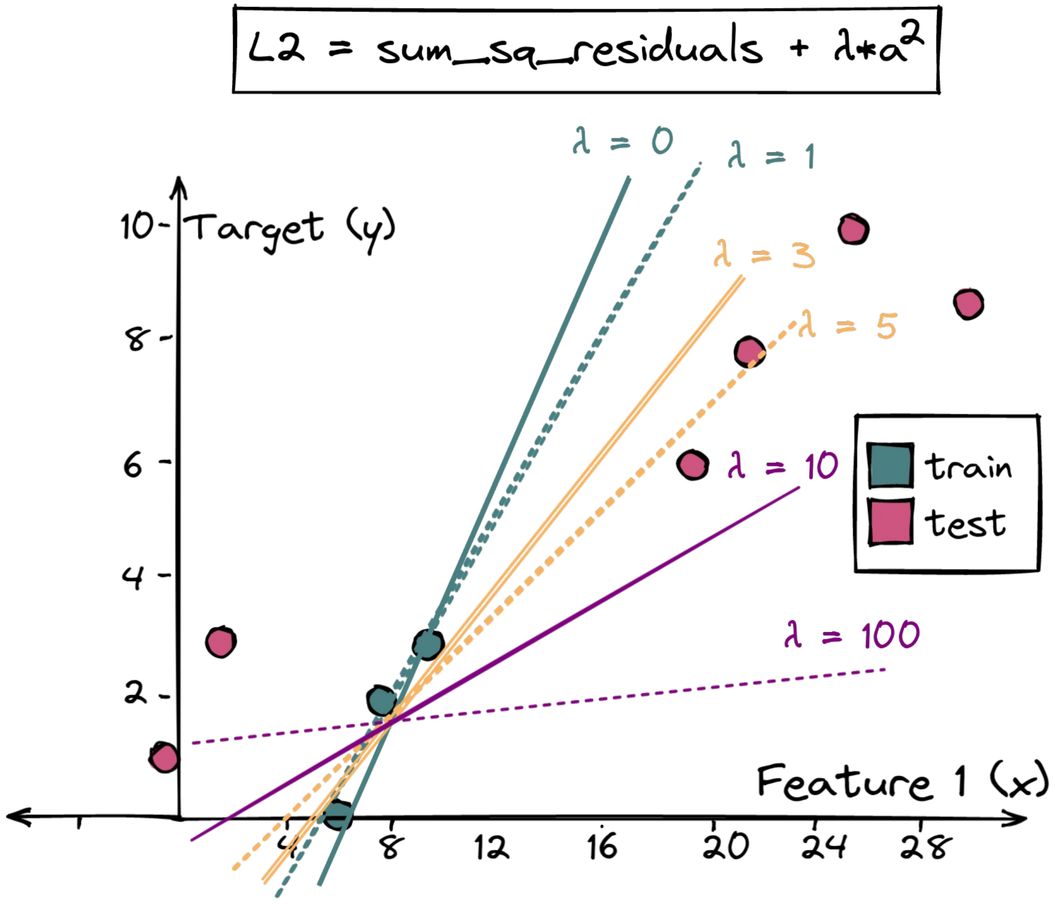
minimizing (sum of squared residuals + $\lambda * a^2$)
λ - lambda

Ridge (L2) Regularization

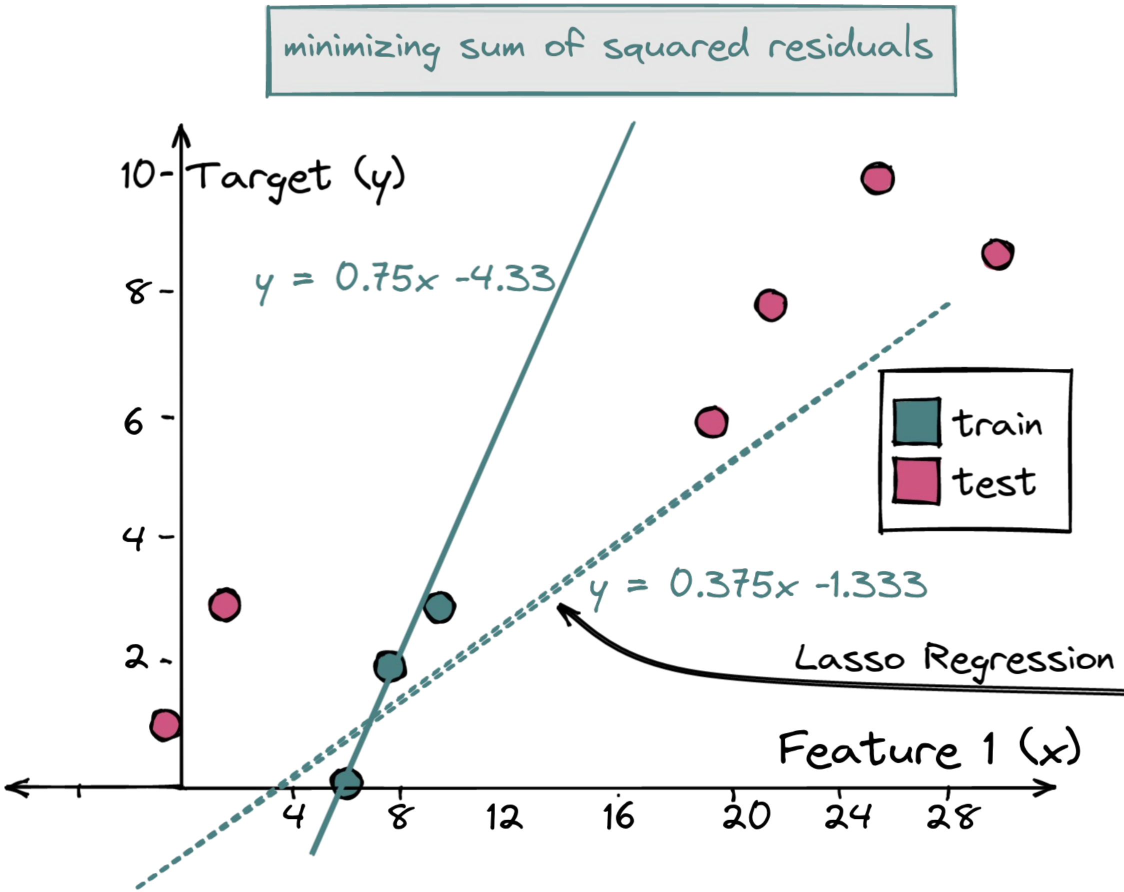
Depending on the value of λ (lambda), the linear fit will vary.

It’s quite impressive, that some of the fitting lines (λ=3,5) describe unseen test data quite well

How do you think you can select λ on practice?



Lasso (L1) Regularization



y = ax + b

L1 = sum_sq_residuals + |λ*a|

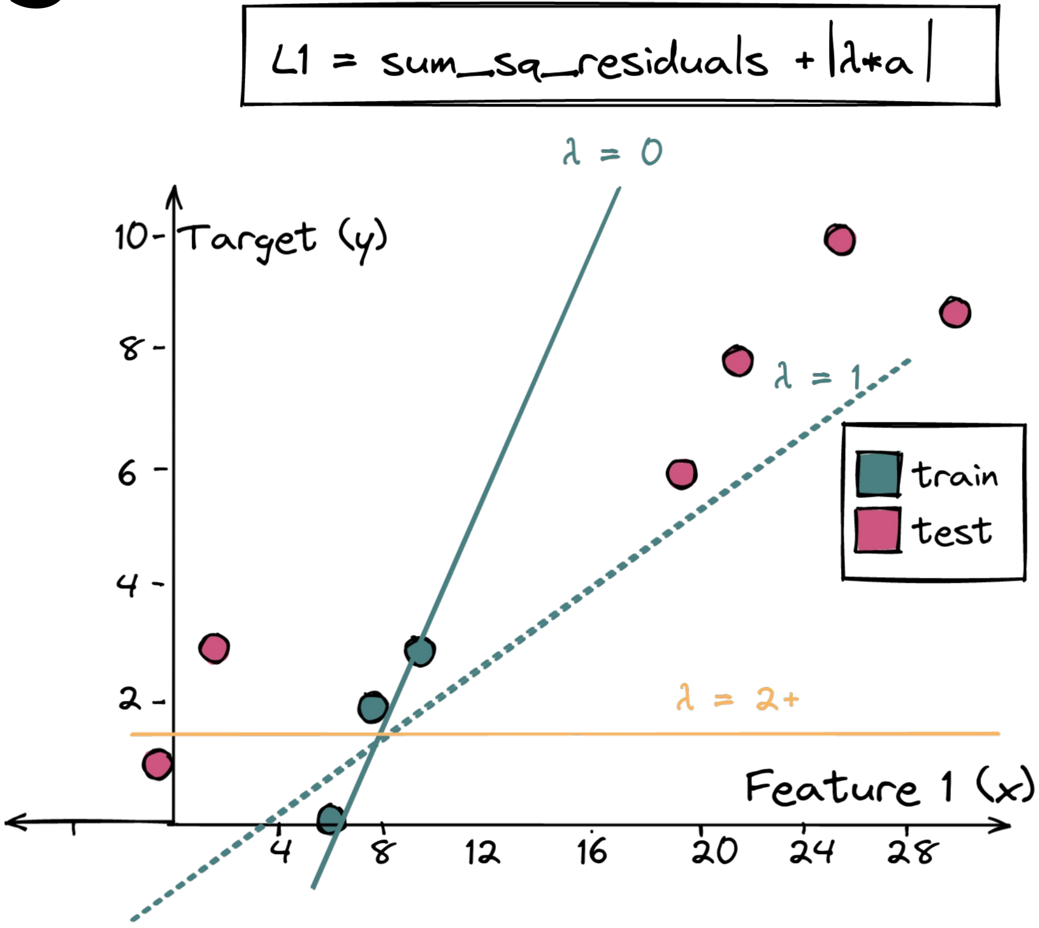
minimizing (sum of squared residuals + |λ*a|)
λ - lambda

Lasso (L1) Regularization

Depending on the value of λ (lambda), the linear fit will also vary similar to Ridge.

Compared to Ridge, the fitting line decrease it's slope much faster

Lasso can actually represent target as constant, whereas Ridge does that asymptotically.



Regularization Fuzz

Multiple linear regression:

$$\boxed{y = a_1 x + a_2 z + a_3 u + b}$$

$$L1 = sum\_sq\_residuals + \lambda * (|a_1| + |a_2| + |a_3| + \ldots)$$

$$L2 = sum\_sq\_residuals + \lambda * (a_1^2 + a_2^2 + a_3^2 + \ldots)$$

Regularization Fuzz

Dubai Taxi example:

$$\boxed{y = ax_1 + az_2 + ay_3 + b}$$

Average waiting time = intercept + a_1*(place) + a_2*(time of day) +
+ a_3*(day of week) + a_4*(your astrological animal) +
+ a_5*(gulf water temperature)

$$L1 = \lambda * (|a_1| + |a_2| + |a_3| + \overset{=0}{\cancel{|a_4|}} + \overset{=0}{\cancel{|a_5|}})$$

Great to use when some variables are useless

$$L2 = \lambda * (a_1^2 + a_2^2 + a_3^2 + \overset{\simeq 0}{a_4^2} + \overset{\simeq 0}{a_5^2})$$

Great to use when most variables are useful