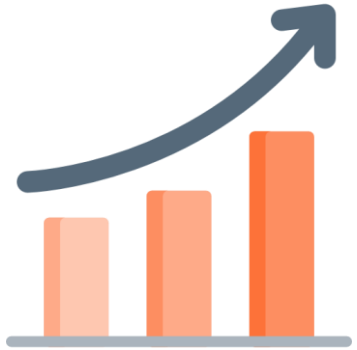
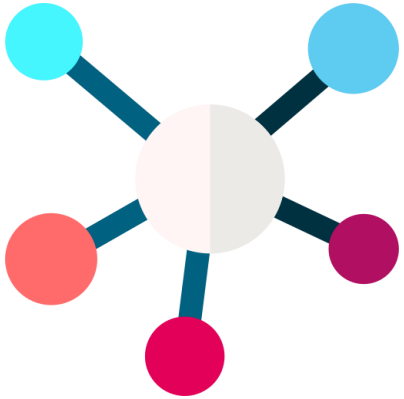
Building Regression Models

Types of Machine Learning Problems

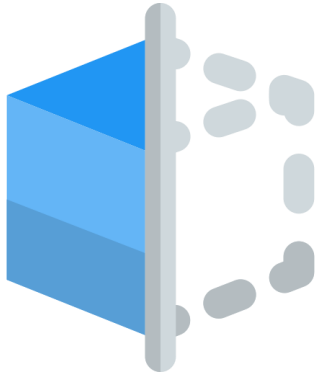
Classification



Regression



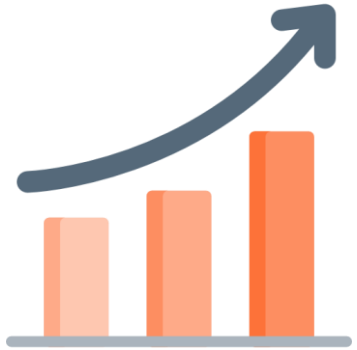
Clustering



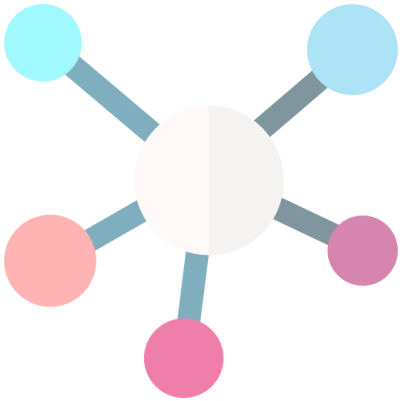
Dimensionality reduction

Types of Machine Learning Problems

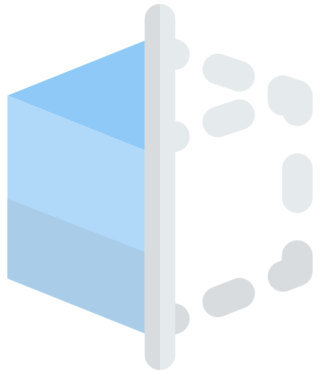
Classification



Regression

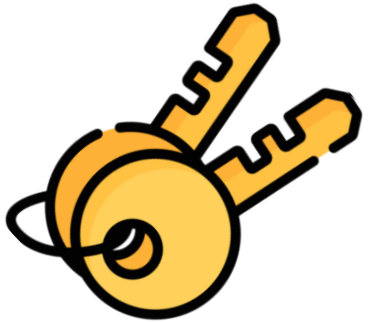


Clustering



Dimensionality reduction

X Causes Y



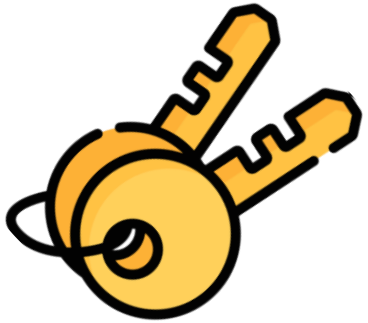
Cause

Independent variable

Effect

Dependent variable

X Causes Y



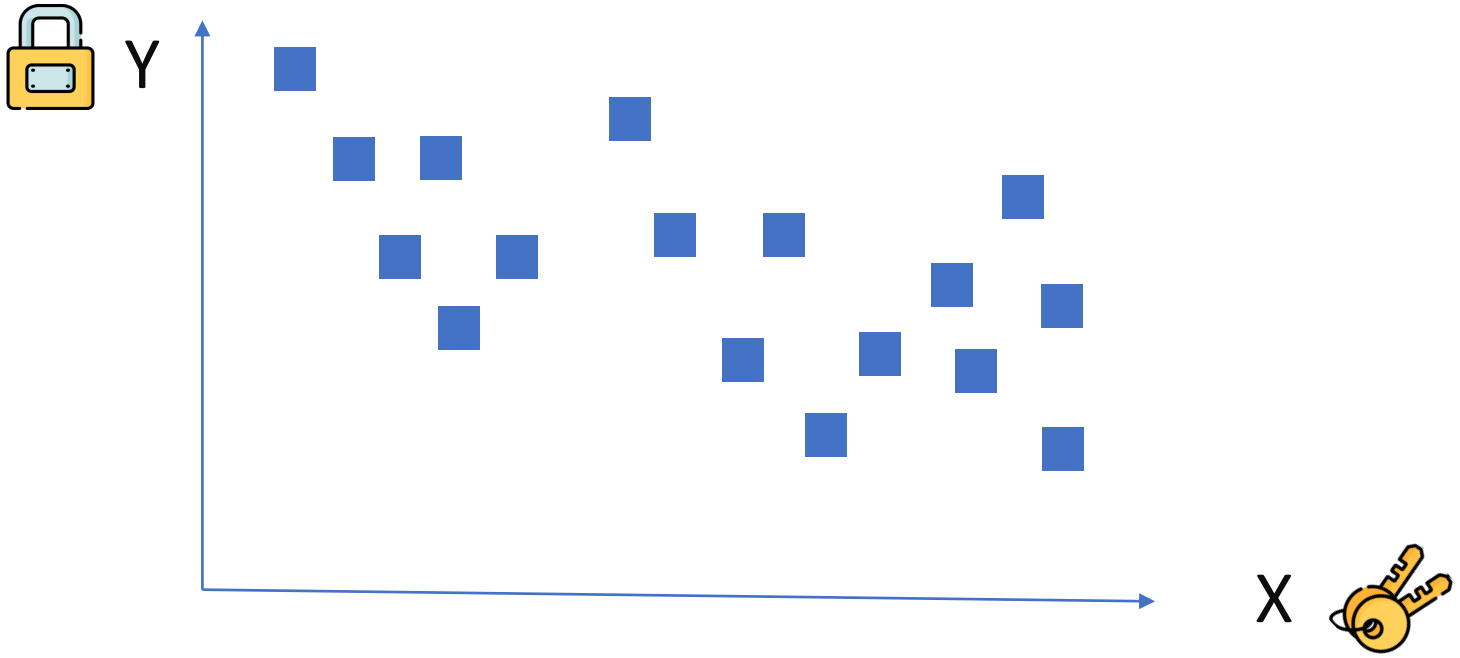
Cause

Explanatory variable

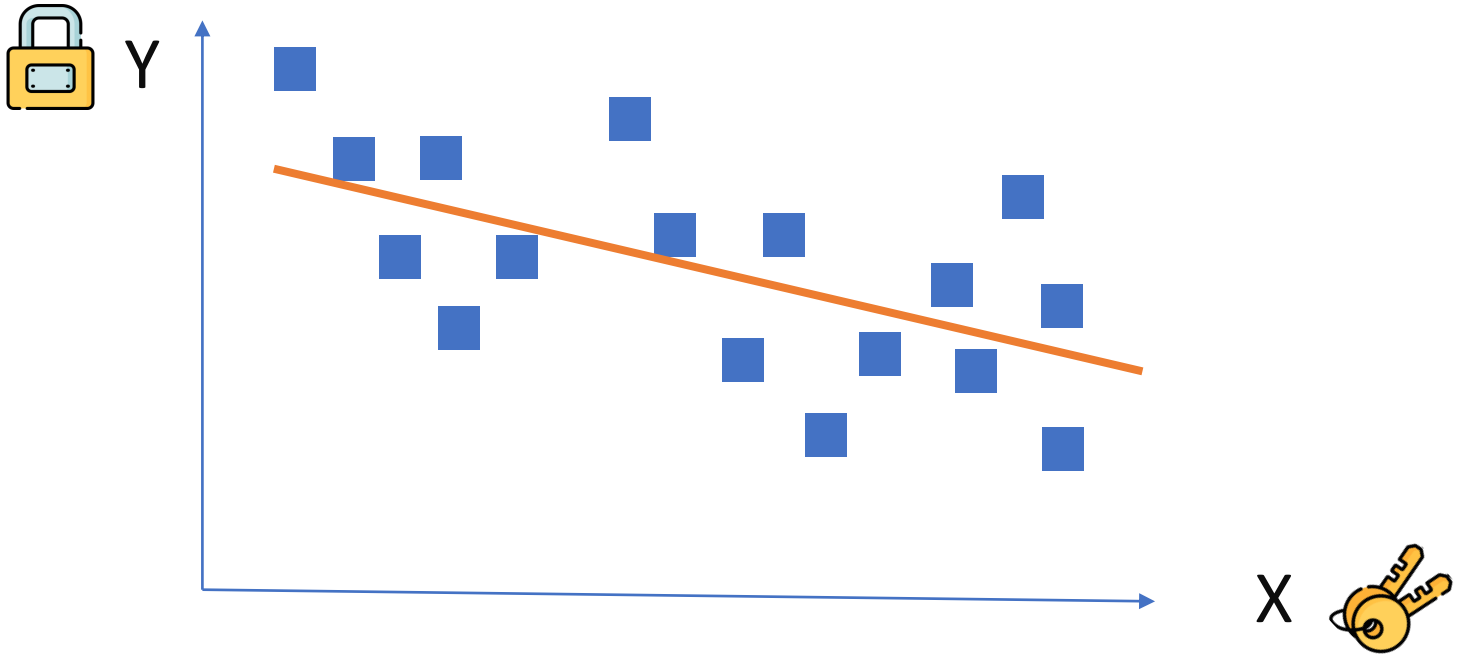
Effect

Dependent variable

The “Best” Regression Line

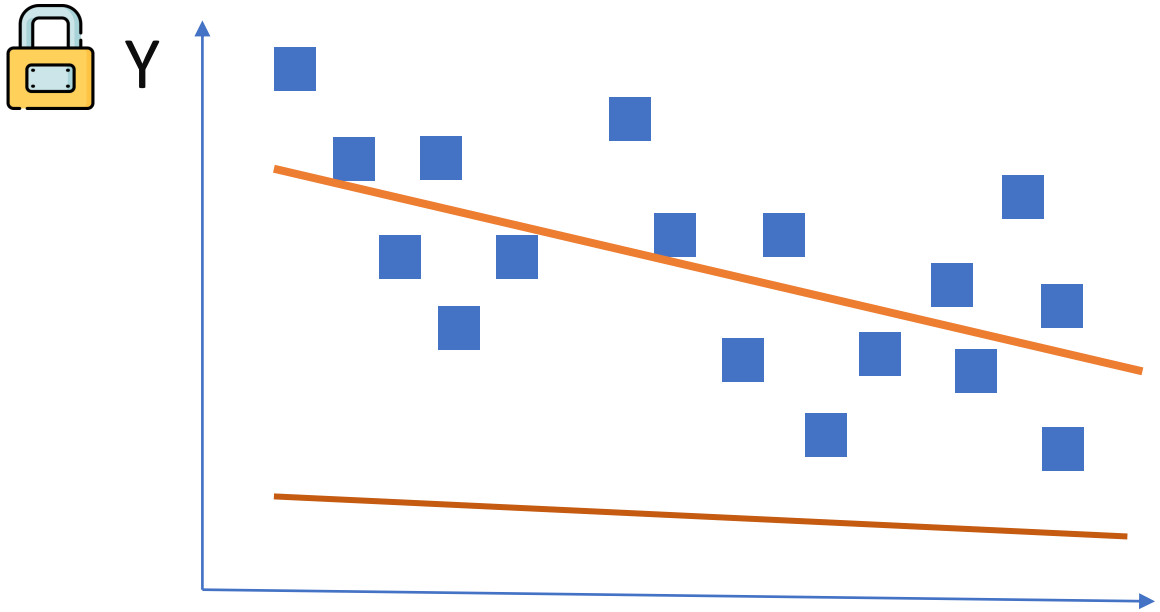


The “Best” Regression Line



Linear regression involves finding the “best fit” line

The “Best” Regression Line

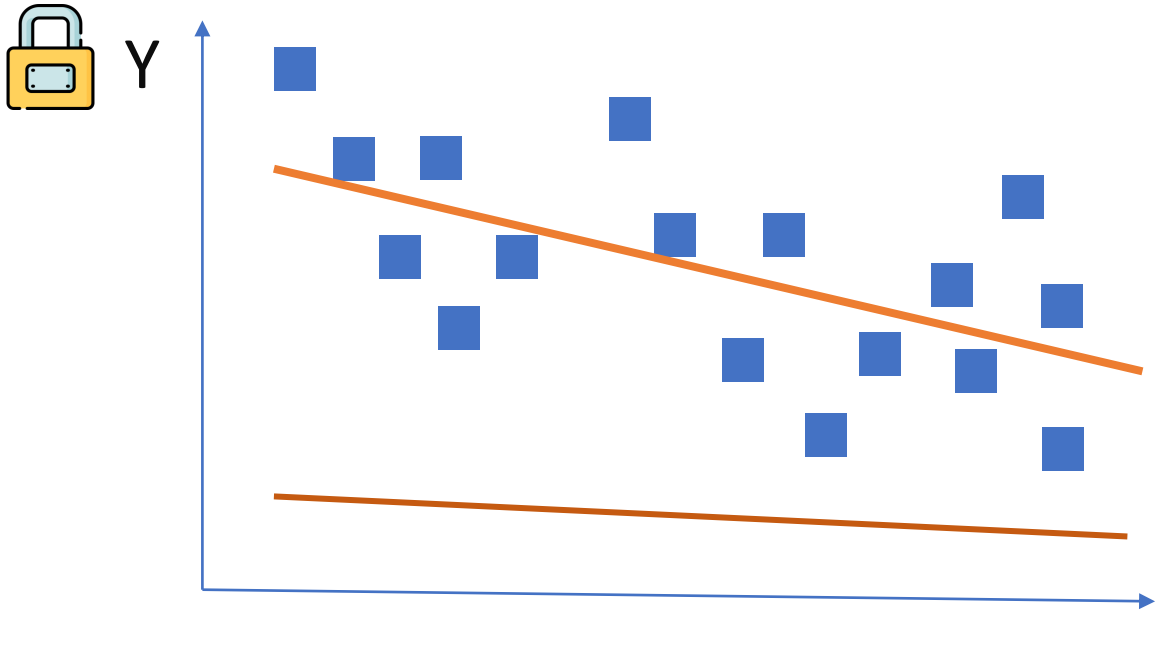


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

Let’s compare two lines, Line 1 and Line 2

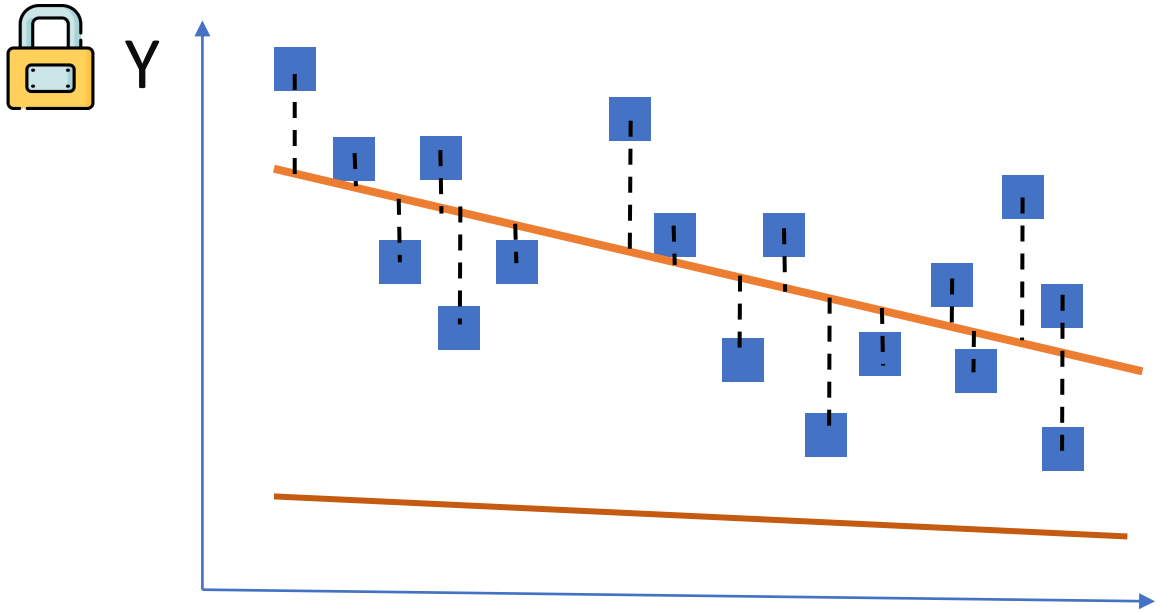
Minimizing Least Square Error



Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

Minimizing Least Square Error

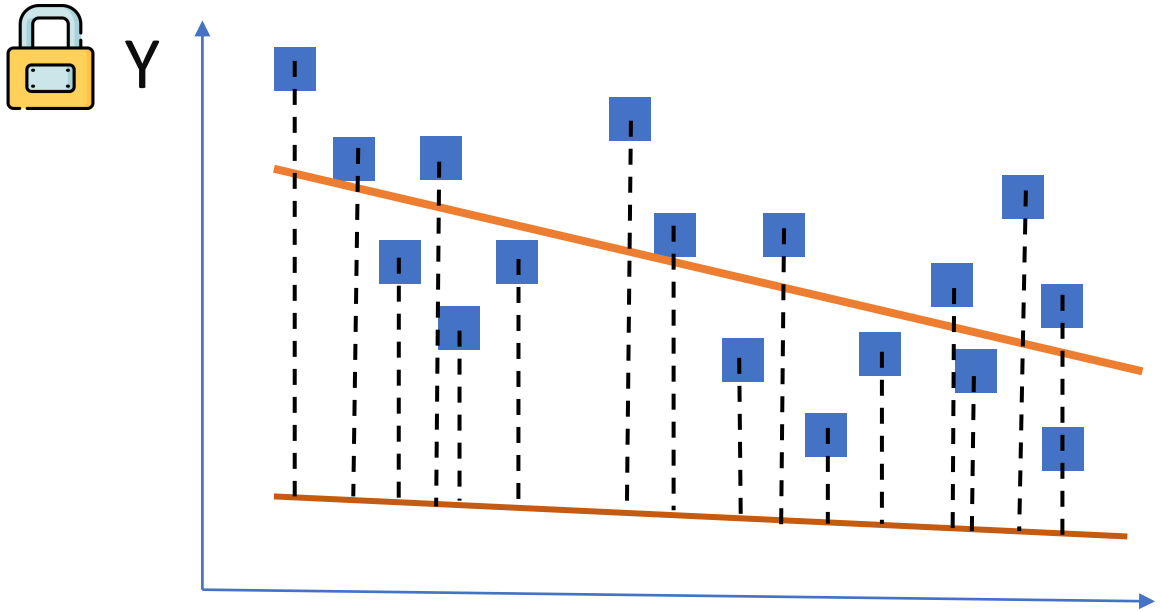


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

Drop vertical lines from each point to the lines 1 and 2

Minimizing Least Square Error

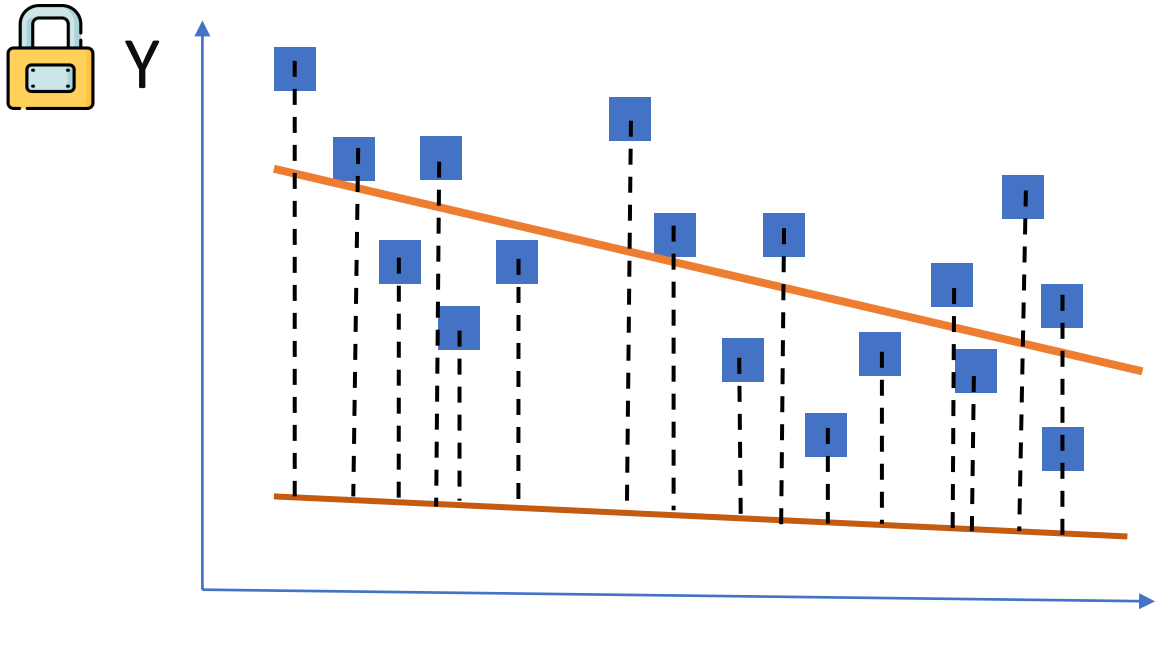


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

Drop vertical lines from each point to the lines 1 and 2

Minimizing Least Square Error

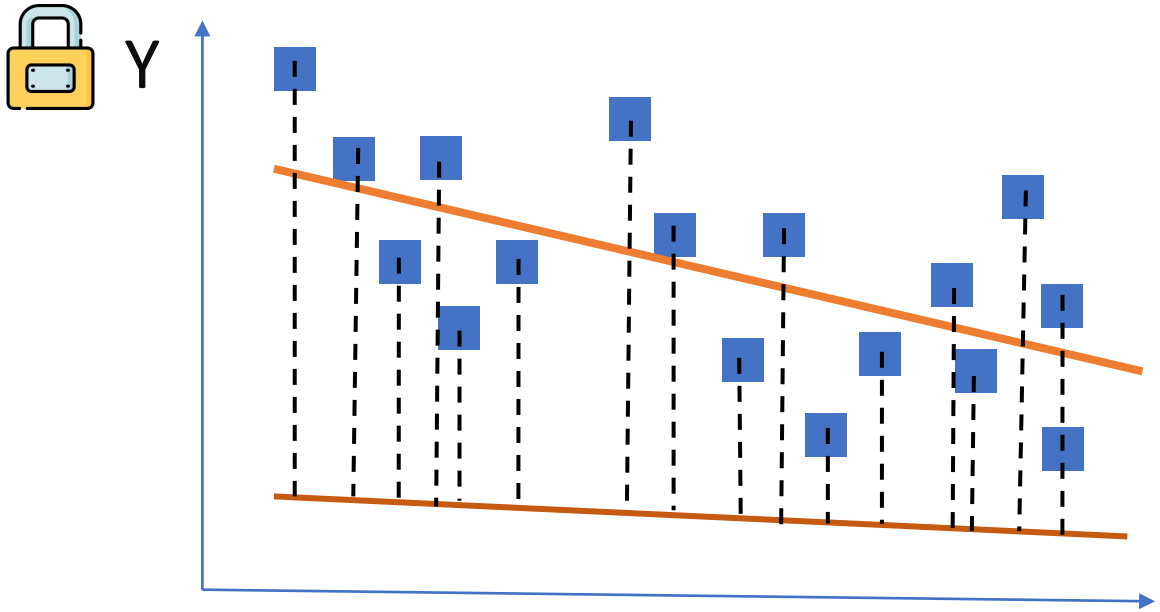


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

The “best fit” line is the one where the sum of the squares of the lengths of these dotted lines are minimum

Minimizing Least Square Error

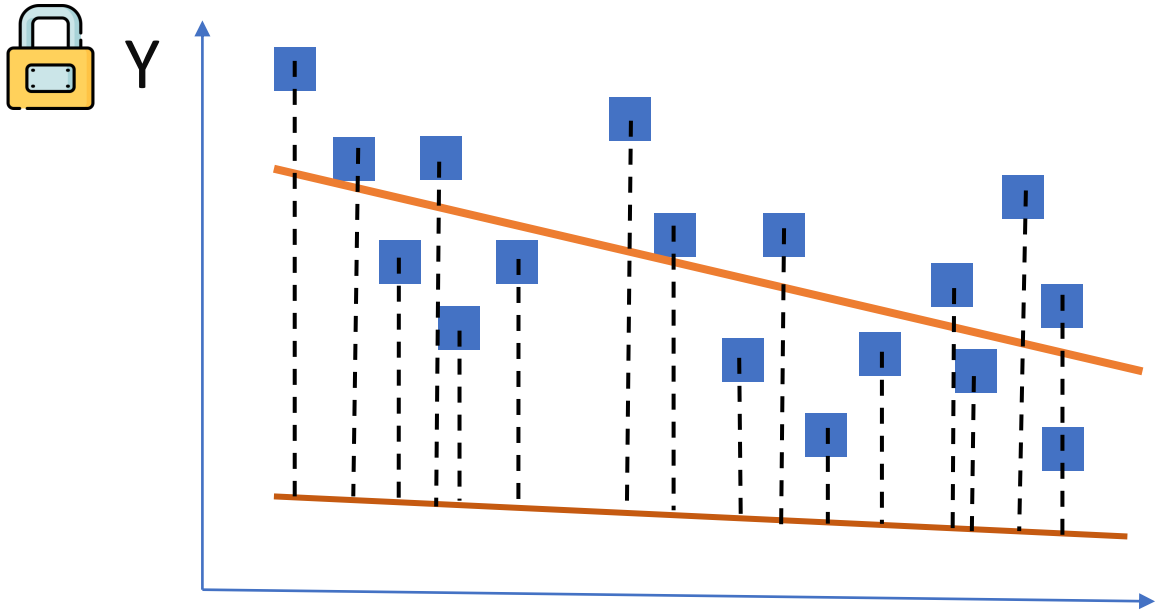


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

The “best fit” line is the one where the sum of the squares of the lengths of these dotted lines are minimum

Minimizing Least Square Error

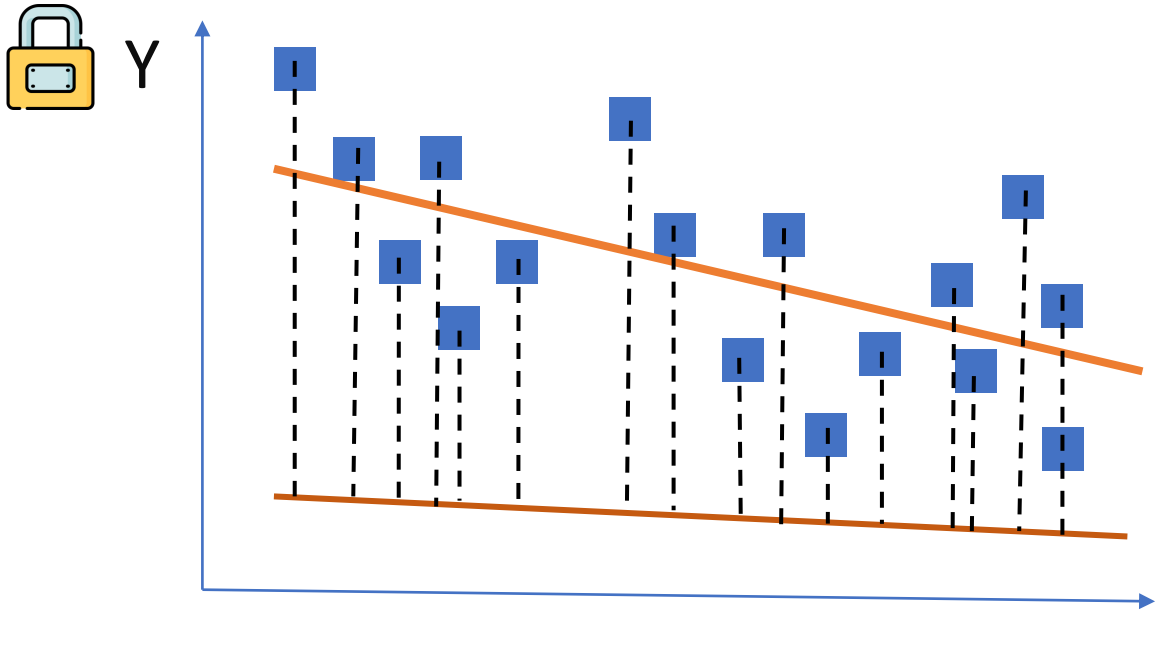


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

The “best fit” line is the one where the sum of the squares of the lengths of these dotted lines are minimum

Minimizing Least Square Error

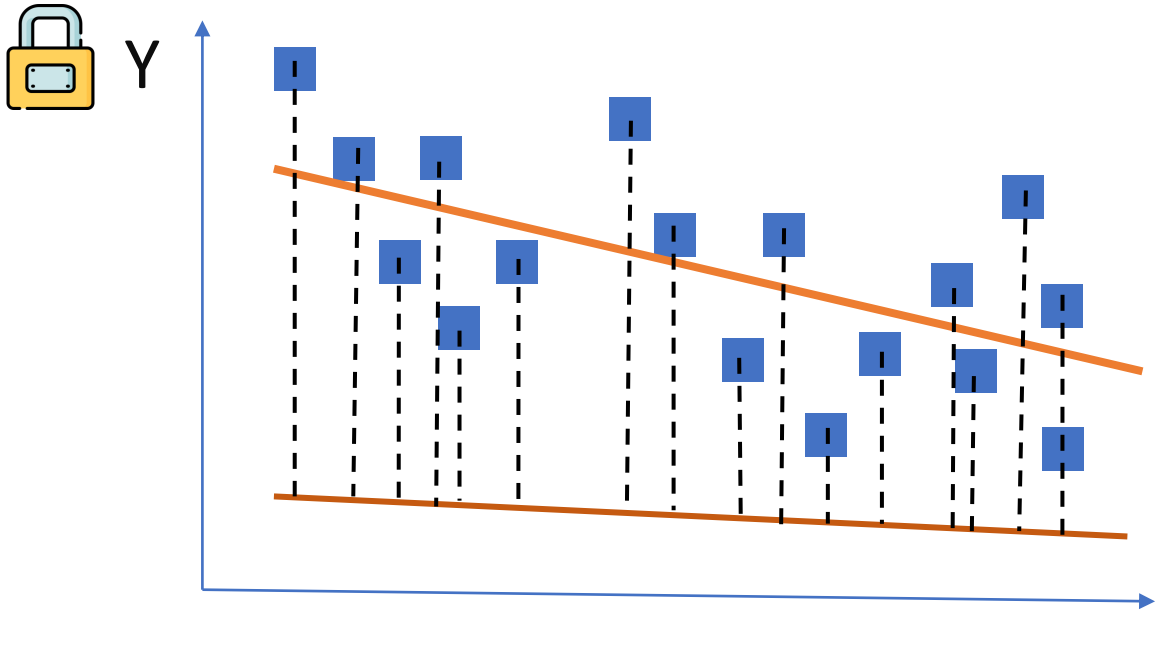


Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

The “best fit” line is the one where the sum of the squares of the lengths of the errors are minimum

Minimizing Least Square Error



Line 1: $y = A_1 + B_1x$

Line 2: $y = A_2 + B_2x$

The “best fit” line is the one where the sum of the squares of the lengths of the errors are minimum