x (feature)

$\hat{Y} = \beta_0 + \beta_1 \cdot x$

x = hours studied = 4

Input value given to model

y (real value)

$error = y - \hat{Y}$

real exam score = 70

Actual correct answer

$\hat{y}$ (ŷ-hat, prediction)

$\hat{Y} = 40 + 5\text{-}4 = 60$

predicted exam score = 60

Model's guess

β or w (weight)

$\hat{Y} = \beta_1 \cdot x$

$\beta_1 = 5$ means x is important

How strong the feature *affects output*

β_0 (bias / intercept)

$\hat{Y} = \beta_0 + \beta_1 \cdot x$

$\beta_0 = 40$ (starting score)

Value when $x = 0$

n (number of samples)

$MSE = \frac{1}{n} \sum (y - \hat{Y})^2$

n = 100 students

Total data points

$\sum$ (sum)

$\sum (y - \hat{Y})^2$

$4 + 9 + 16 = 29$

Add all values

X (input matrix)

$\hat{Y} = X\beta$

X = table of inputs

All input data together

J or L (loss)

$J = (y - \hat{Y})^2$

$(70 - 60)^2 = 100$

How wrong the model is

λ (lambda, regularization)

$Loss = MSE + \lambda\beta^2$

λ = 0.1 controls weight size

Prevents overfitting

α (learning rate)

$\tilde{\beta} = \beta - \alpha \cdot error$

α = 0.01 small step

How fast model learns

i (index)

$\tilde{y}_i - \tilde{y}_i$

$i = 3 \rightarrow$ third data row

Refers to one data point.

Symbol	Meaning	Equation	Example
x	Feature / Input	$\hat{y} = \beta_0 + \beta_1 x$	$x =$ hours studied $= 4$
y	Actual Value	$error = y - \hat{y}$	Real exam score = 70
$\hat{y}$	Prediction	$\hat{y} = 40 + 5 \cdot 4 = 60$	Predicted score = 60
β or w	Weight	$\hat{y} = \beta_1 x$	$\beta_1 = 5 \rightarrow$ strong effect
β_0	Bias / Intercept	$\hat{y} = \beta_0 + \beta_1 x$	$\beta_0 = 40 \rightarrow$ base score
n	Number of Samples	$MSE = \frac{1}{n}\sum(y - \hat{y})^2$	$n = 100$students
Σ	Summation	$\sum(y - \hat{y})^2$	$4 + 9 + 16 = 29$
X	Input Matrix	$\hat{y} = X\beta$	Table of all inputs
J or L	Loss Function	$J = (y - \hat{y})^2$	$(70 - 60)^2 = 100$
λ	Regularization	$Loss = MSE + \lambda\beta^2$	$\lambda = 0.1$
α	Learning Rate	$\beta = \beta - \alpha \cdot error$	$\alpha = 0.01$
i	Index	$y_i - \hat{y}_i$	$i = 3 \rightarrow$ third data point