☆ Linear Regression

① Data

x_1	x_2	y
0.2	-1.3	3.1
-3.8	2.6	-0.2
-1.2	1.3	8.3

② Prediction eqⁿ:

$$\rightarrow \underline{y_{pred}} = w_1 x_1 + w_2 x_2 + b$$

Optimize them!

$\left.\begin{matrix} w_1 \\ w_2 \end{matrix}\right\}$ model weights

$b \}$ model bias

model params = Total weights + Total biases.

3 Cost Function

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

↓

High → Model needs more training

low → Good Sign! Model needs less training

Converging → We can stop the training

y	$\hat{y}$
0.3	100.8
-1.8	26.3
2.6	-85.6

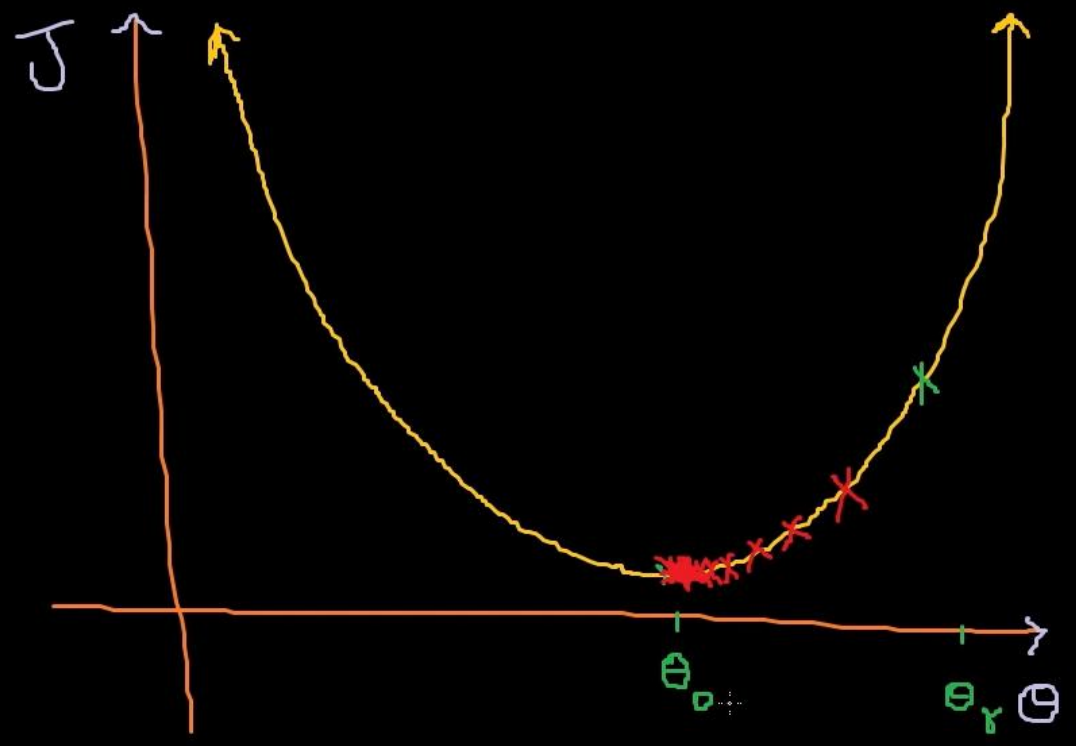
(4) Optimisation – Gradient Descent Algorithm

$$\theta = \theta - \alpha \boxed{\frac{\partial J}{\partial \theta}}$$

learning rate (0.001)

gradient at given θ

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



☆ Derivatives

→ $y = ax^n$

→ $\frac{dy}{dx} = n \cdot ax^{n-1}$

→ $y = 3x^2$

$\therefore \frac{dy}{dx} = 6x^1$

☆ Chain Rule

→ $y = k(ax+b)^n$ → $\frac{dy}{dx} = nk(ax+b)^{n-1} \cdot a$

→ $y = k\,f(x)^n$

$\therefore \frac{dy}{dx} = n \cdot k\,f(x)^{n-1} \cdot f'(x)$

→ $y = -2(3x+8)^2$

→ $\frac{dy}{dx} = -4(3x+8)^1 \cdot (3) = -12(3x+8)$

★ **Partial derivatives**

$$y = x^2 t^2 + 3xt^3 - 3t + 2x$$

$$\frac{\partial y}{\partial x} = 2xt^2 + 3t^3 + 2$$

$$\frac{\partial y}{\partial t} = 2tx^2 + 9xt^2 - 3$$

★ Partial derivations for MSE (Cost funcⁿ)

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - (Wx_i + b)\right)^2$$

① $$\frac{\partial J}{\partial W} = \frac{1}{N}\sum_{i=1}^{N}\left[2\left(y_i - (Wx_i + b)\right) \cdot (-x_i)\right] = \frac{1}{N}\sum_{i=1}^{N}\left[-2x_i\left(y_i - \hat{y}_i\right)\right]$$

② $$\frac{\partial J}{\partial b} = \frac{1}{N}\sum_{i=1}^{N}\left[2\left(y_i - (Wx_i + b)\right) \cdot (-1)\right] = \frac{1}{N}\sum_{i=1}^{N}\left[-2\left(y_i - \hat{y}_i\right)\right]$$