Conventional ML

- Prepare the data
- Train model from training data to estimate models Parameters
- Stores Models in Suitable forms
- Predict unseen scoring instance

Instance Based Learning

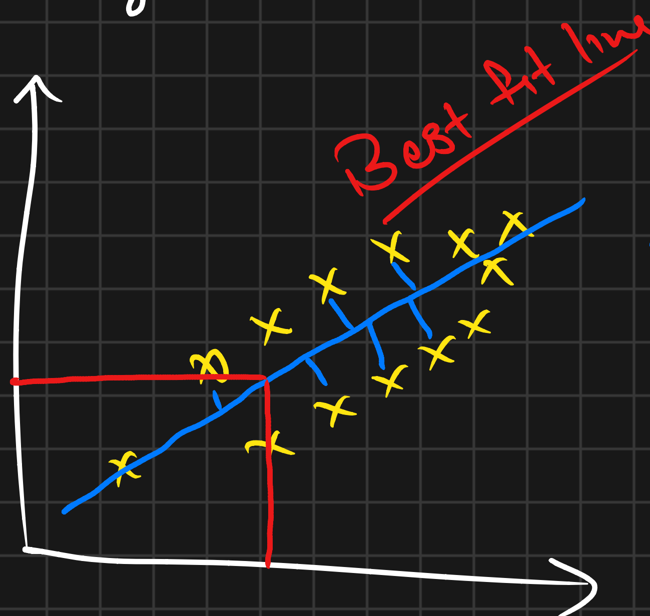
- Do not train model Pattern discovery postponed until scoring query
- There is no model to store

Simple Linear Regression

Supervised ML:

weight	height
74	170
80	180
75	175.5cm

new weight → **Model** → height

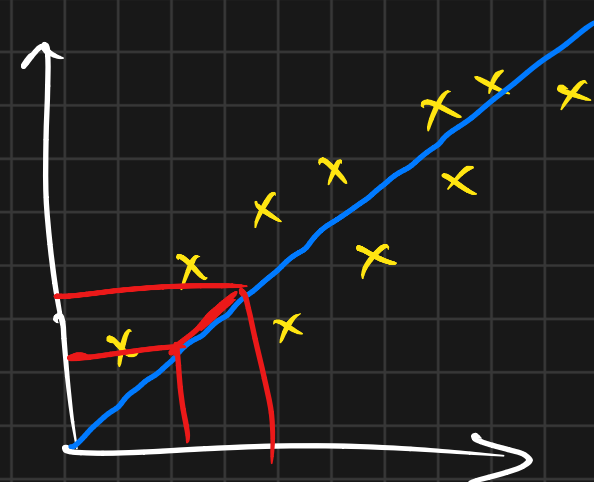


error sum should be minimum

$y = mx + c$

$y = \beta_0 + \beta_1 x$

$h_\theta(x) = \theta_0 + \theta_1 x$



$h_\theta(x) = \theta_0 + \theta_1 x$

θ_0 = Intercept

θ_1 = slope or coffeicent

$h_\theta(x)$ = Predicted Point

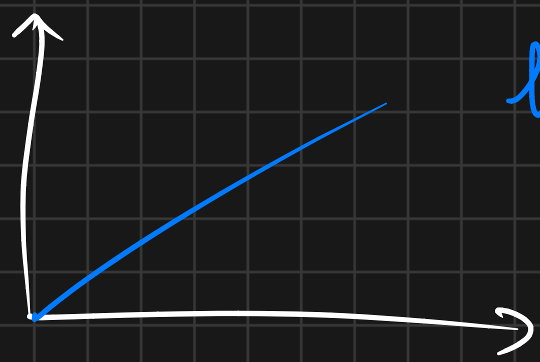
Cost function

$$J(\theta, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2$$

• mean squared error

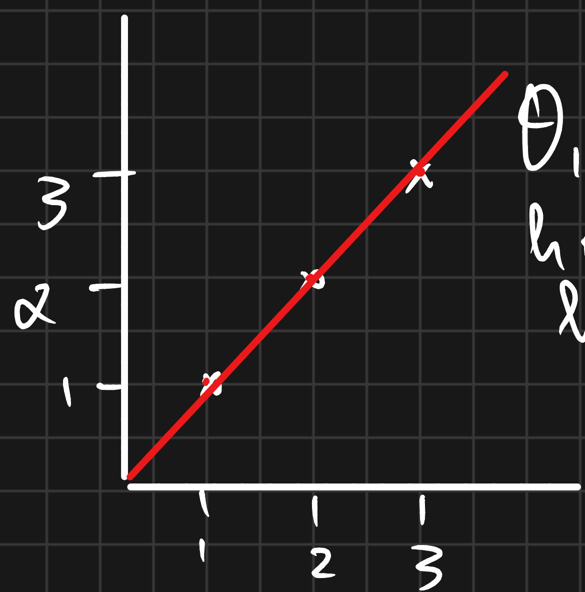
minimize $J(\theta, \theta_1) = \frac{1}{2m} \sum_{i=1}^{M} \left(h_\theta(x)^{(i)} - y^{(i)} \right)^2$

① $h_\theta(x) = \theta_0 + \theta_1 x$ $\boxed{\theta_0 = 0}$



$h_\theta(x) = \theta_1 x$

x	y
1	1
2	2
3	3



$\theta_1 = 1$

$h_\theta(x) = 1$ $x=1$

$h_\theta(x) = 2$ $x=2$

$h_\theta(x) = 3$ $x=3$

$$J(\theta_0, \theta_1) = \frac{1}{2 \times 3} \sum_{i=1}^{3} \left[(1-1)^2 + (2-2)^2 + (3-3)^2 \right]$$

$\boxed{J = 0}$

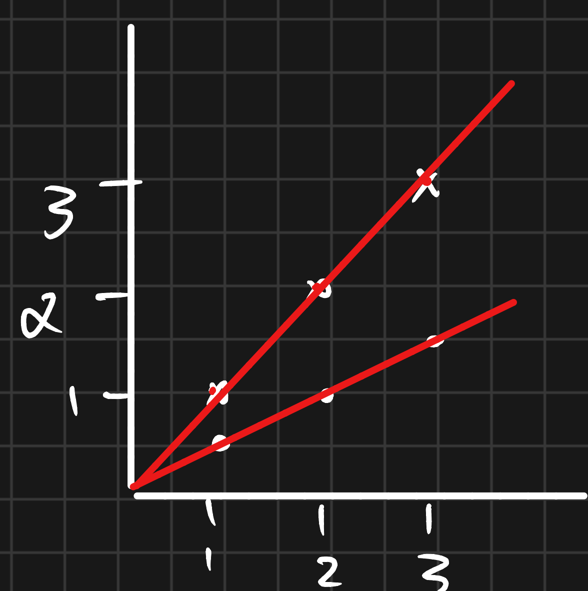
let $h_\theta(x) = \theta_1 x$

$h_\theta(x) = 0.5$ $x = 1$

$\theta_1 = 0.5$

$\theta_2 = 1$

$\theta_3 = 1.5$



$$\frac{1}{2 \times 3} \sum_{i=1}^{3} \left[(0.5-1)^2 + (2-1)^2 + (1.5-3)^2 \right]$$

$= .58$

$J(\theta_1) = 0$ ←

$J\theta_0, J\theta_1, J\theta_2$

gradient decent

Global Minima

Convergence Algorithm

Optimze the changes of θ_1 value

Repeat until Convergence

{

θ_1 value change efficiently

$$\left\{ \theta_J = \theta_J - \alpha \frac{\partial J(\theta_J)}{\partial \theta_J} \right.$$

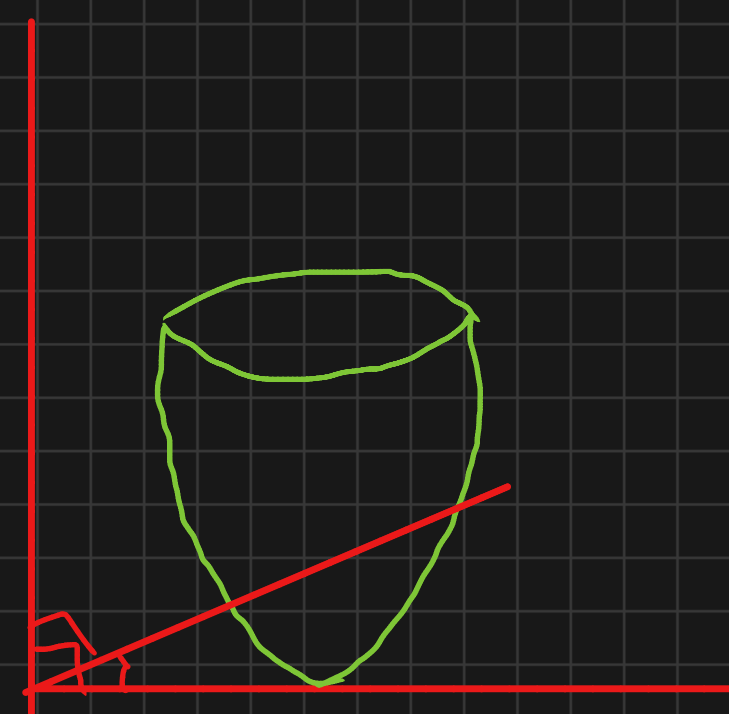
}

$$\left\{ \begin{array}{l} \theta_J = \theta_J - \alpha(-ve) \\ \theta_J = \theta_J + (+ve) \end{array} \right\}$$

$$\left\{ \begin{array}{l} \theta_J = \theta_J - \alpha(+ve) \\ \theta_J = \theta_J - (+ve) \end{array} \right.$$

α (alpha): learning Rate if it is very small it will keep on iterating if α is too big then it will vary too much

Gradient Descent



$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{i} - y^{(i)}\right)^2$$

$$J = 0 \Rightarrow \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) =$$

$$\frac{\partial}{\partial \theta_0} \frac{1}{2m} \left[\sum_{i=1}^{m} \left(h_\theta(x)^{i} - y^{i}\right)^2 \right]$$