Linear Regression

12 March 2023 12:54

Machine Learning

Supervised — labelled data → Regression, Classification

Unsupervised

X	y (labels / Dependent Variable)
1 1 1	a 2.5
	b 2.6
	c

ee

SUPERVISED

1. Advantages

- Easy to train
- Better results due to training of past data
- New samples can easily be trained

2. Disadvantages

- Classifying / Processing / Collecting data is a tedious task
- Labelling is challenging
- Lot of computation time
- Complex tasks → fails

LINEAR REGRESSION

1. Supervised
2. Regression

Eg: House Price

Dataset

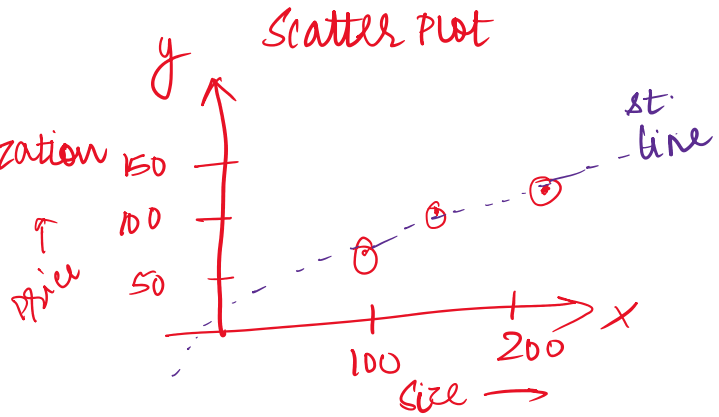
x : size of house

y : price

x (size)	y (price)
200	126
150	105
100	70

data pts / instances

Visualization



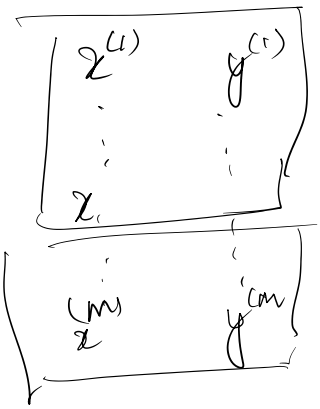
$[x^{(1)}\ y^{(1)}]\ [x^{(2)}\ y^{(2)}]$ $\quad [x^{(m)}\ y^{(m)}]$ → data

$[x^{(1)}, y^{(1)}], [x^{(2)}, y^{(2)}] \dots [x^{(m)}, y^{(m)}]$ → data

x^i : x variable of i^{th} data instance

y^i : y variable of ----

m : total data instances



Training Data
80-70%

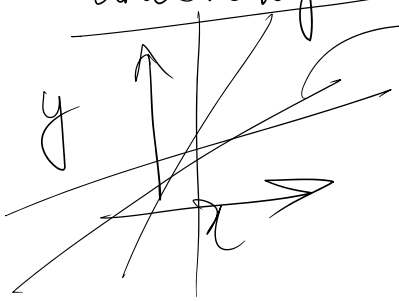
Testing Data
20-30%

Training Data
↓
Learning Algorithm
↓
[h]
↓
Learned o/p f'n
that maps from x to y

$h_\theta(x) = y$
↳ parameters

h
↓
eqn ?

Linear Regression



$y = mx + c$

$h_\theta(x) = \theta_1 x + \theta_2$

y is linearly dependent on x

$\Rightarrow y \propto x$

$\Rightarrow y = mx + c$

θ_1 → Slope
θ_2 → intercept

To find the most optimal value of θ_1 & θ_2 for given data

Gradient Descent Algorithm

To find θ_1 & θ_2

1st Round

Guess : $\theta_1 = 10$, $\theta_2 = 0.5 = 1/2$

$h_\theta(x) \Rightarrow x = 120$

$\rightarrow 120 \times \frac{1}{2} + 10 = 70$

x	y
120	64
90	56
190	94

Train Data

LR error $= [y - f_0(x)]^2 = [64 - 70]^2 = 6^2 = 36$

$\text{error}_1^{(1)} = 36$

$\text{error}_1^{(2)} = [55 - 56]^2 = 1$

$\text{error}_1^{(3)} = [85 - 96]^2 = 11^2 = 121$

subscript = round

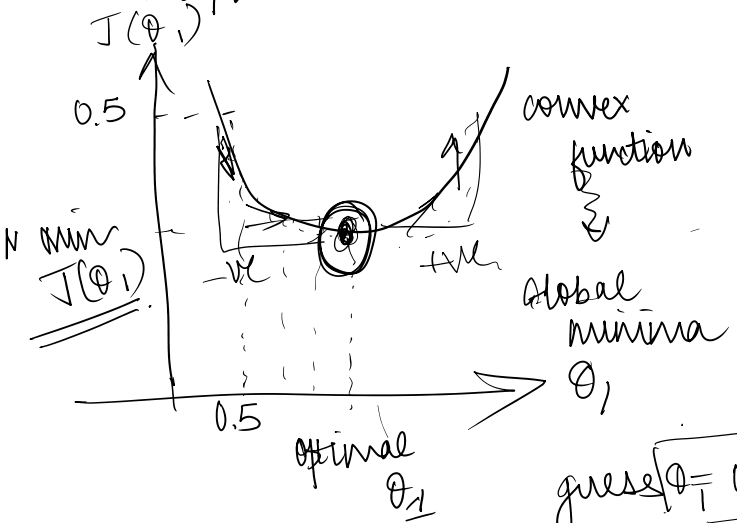
superscript = instance

Total Error in 1st round $= \frac{1}{3} \sum_{j=1}^{3} \text{error}_1^{(j)}$

$= \frac{1}{3} [36 + 1 + 121]$

Mean Square Error → Cost function / loss function

cost/loss should be minimized in every round.



$\theta_1 = \theta_1 - \alpha \left(\frac{\partial J(\theta_1)}{\partial \theta_1} \right)$

learning rate

gradient → slope. → by what amt J will change w.r.t. θ

guess $\theta_1 = 0.5$

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

→ all training

↑ Error

Error ↑

15 46 47 48 50 51 52

Assumption

Min

True