Machine Learning

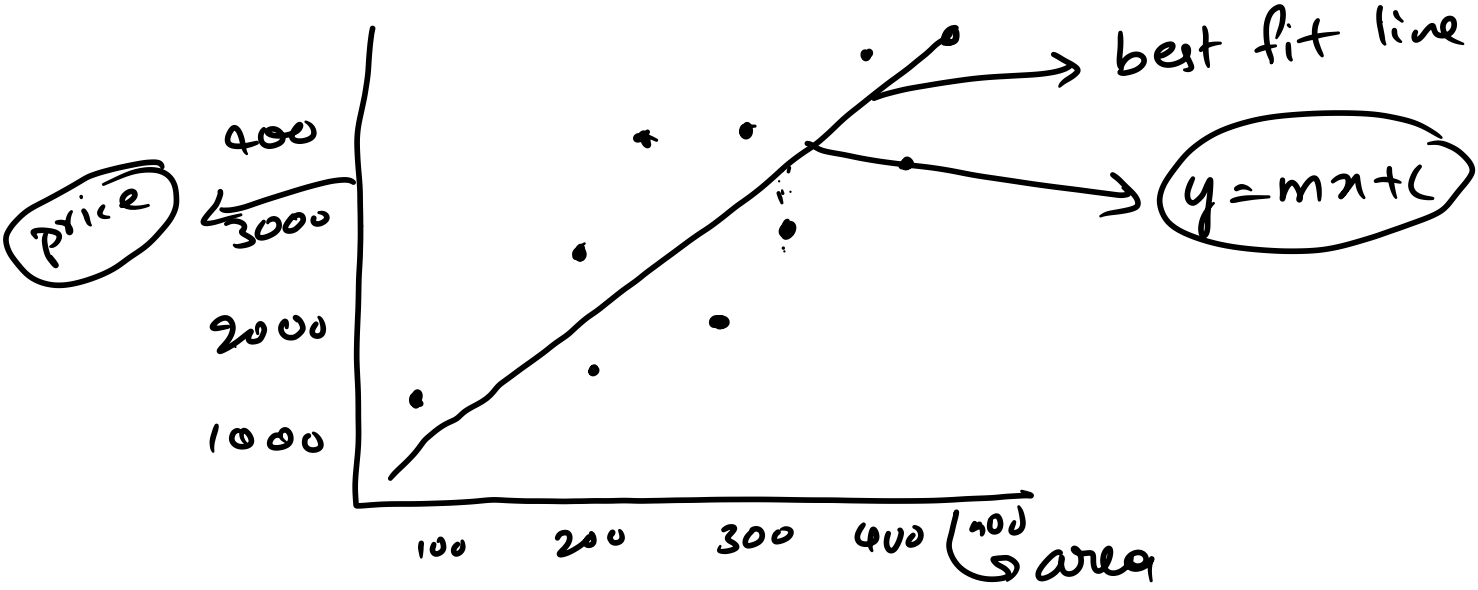
- Linear regression
- Logistic regression
- Decision tree
- SVM (Support Vector Machine)
- Random Forest Algorithm
- Navis Bayes [Probability]

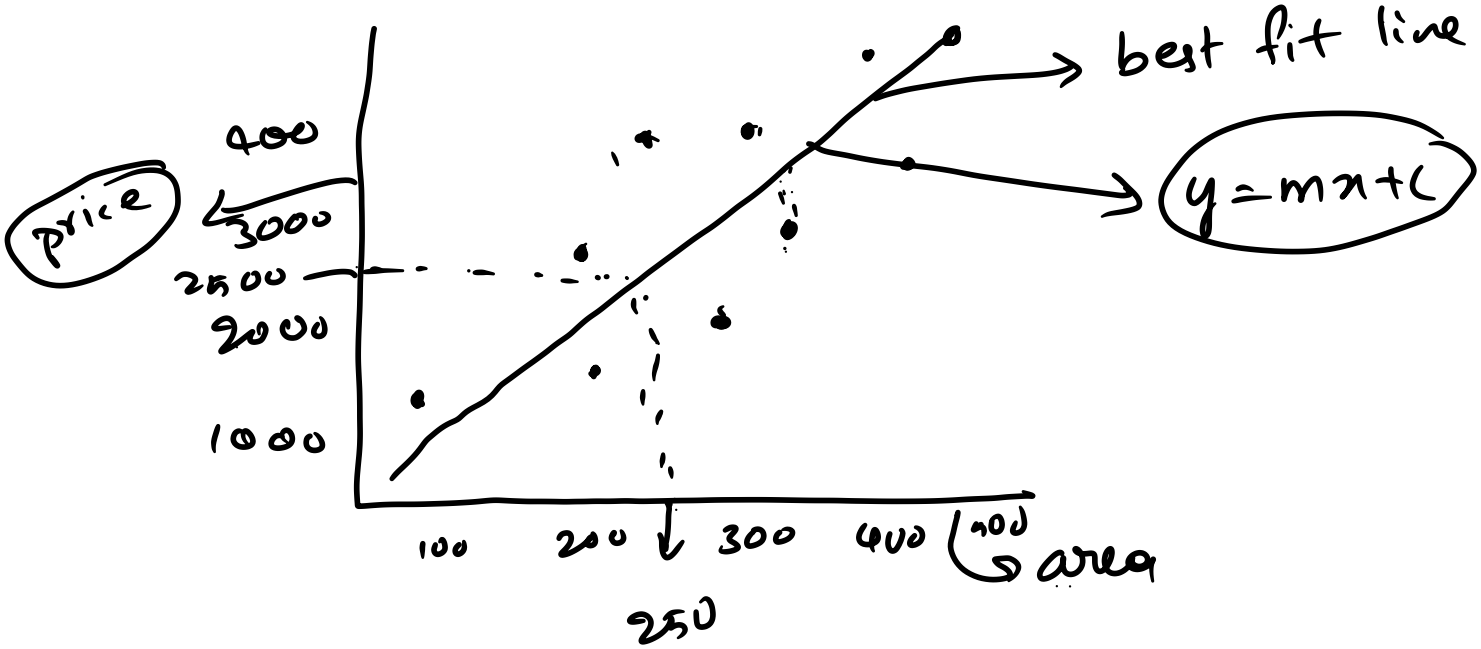
⇒ Linear Regression

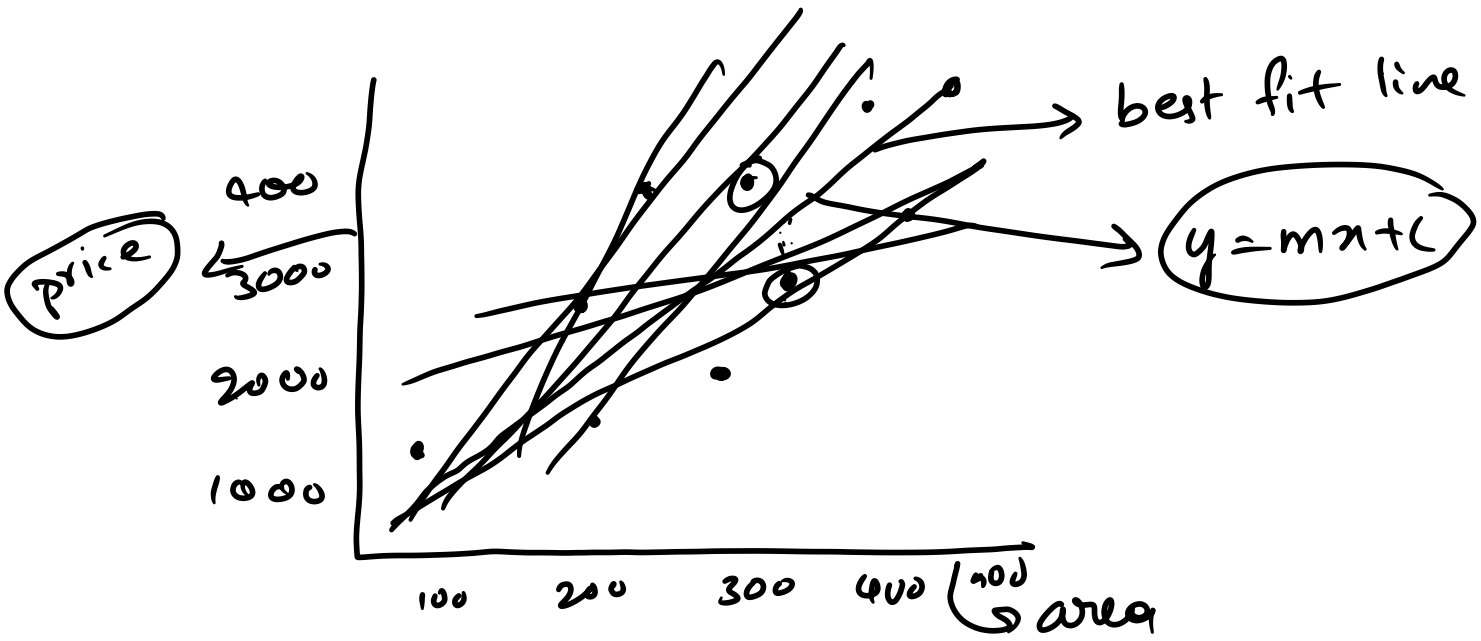
target

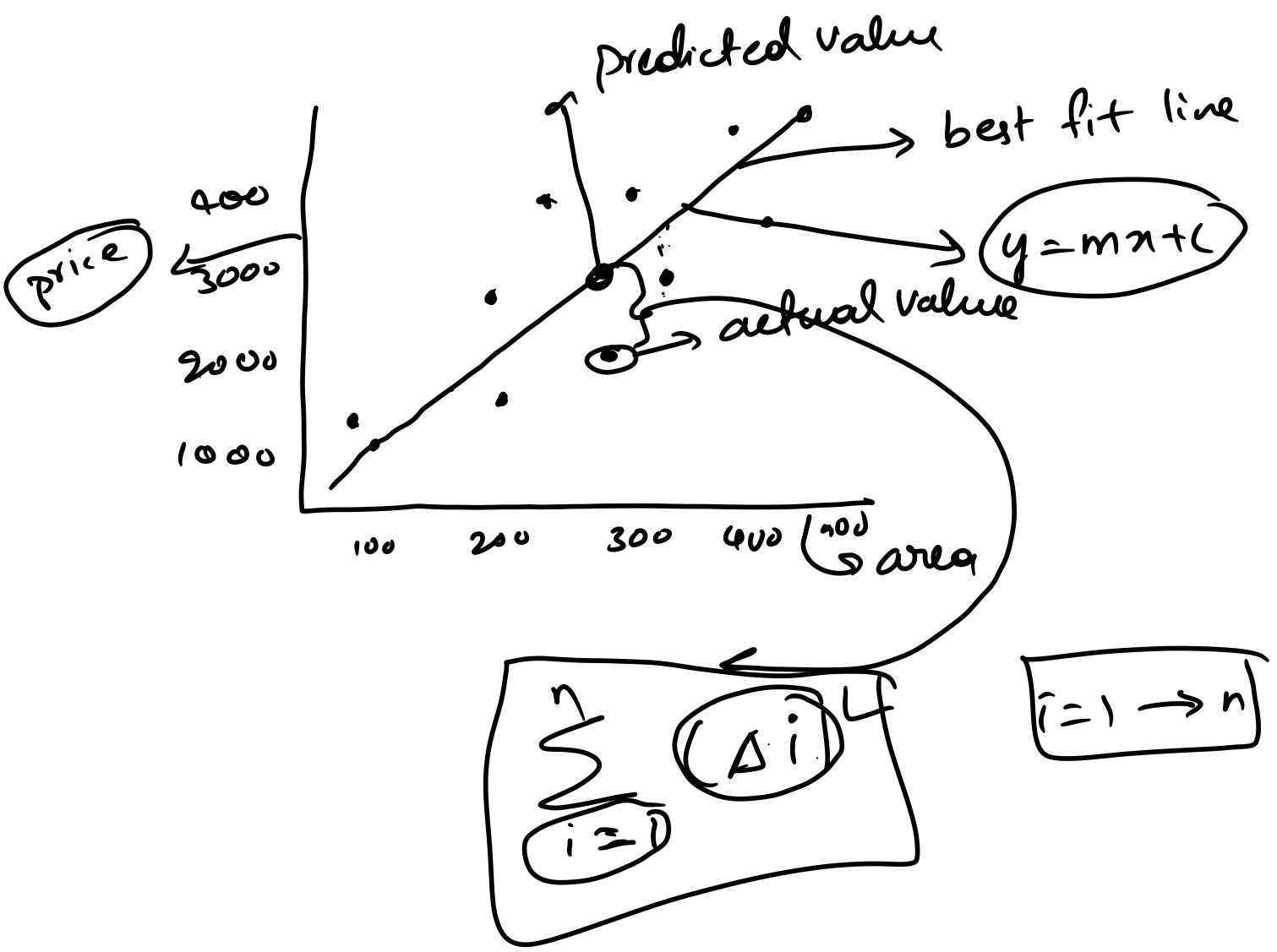
area	price
2600	55000
3200	65000
4000	70000
2000	61000
1500	50000

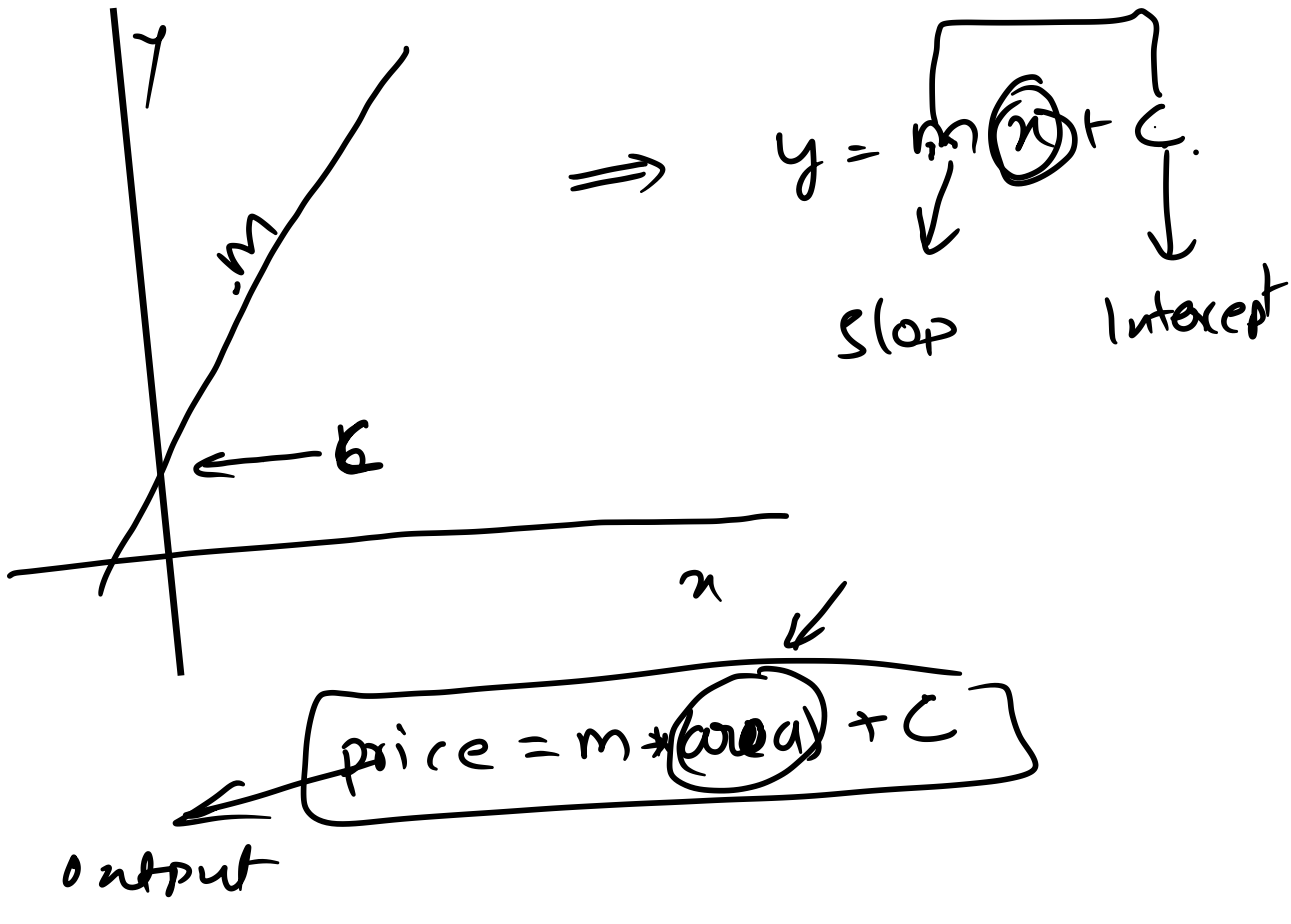
3300 → predict




best fit line
y=mx+c
price
400
3000
2500
2000
1000
100
200
300
400
500
area
250


best fit line
y=mx+c
price
400
3000
2000
1000
100
200
300
400
500
area


predicted value
best fit line
y = mx + c
actual value
price
4000
3000
2000
1000
100
200
300
400
500
area
$\sum_{i=1}^{n} \Delta i$
i = 1 → n


Y
m
c
X
y = mx + c
slop
Intercept
price = m * (area) + C
area
output

Linear Regress with Multiple Variables

(1) multivariate Regression

area	bedrooms	age	price
2600	3	20	550000
3000	4	15	
		.	

3000 sq ft area, 3 bedrooms, 40 years old

price = $m_1 * \text{area} + m_2 \times \text{bedrooms} + m_3 \times \text{age} + c$

price → dependent feature

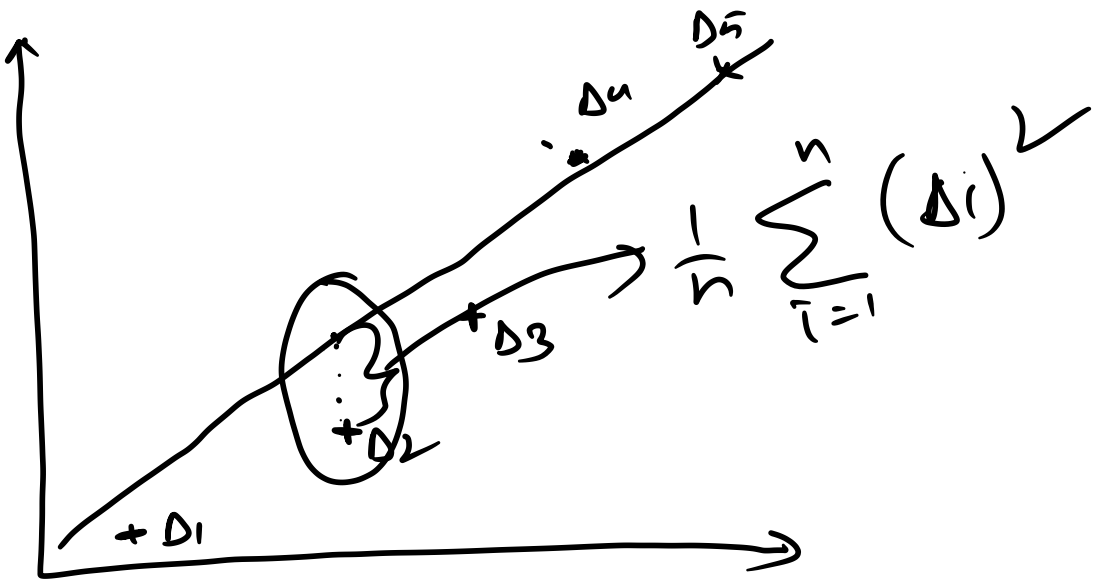
area, bedrooms, age → independ variables

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + c$$

$x = [1, 2, 3, 4, 5]$

predicted $y = [5, 7, 9, 11, 13]$

$y = 2x + 3$ (1)



mean Squared Error

$$mse = \frac{1}{n}\sum_{i=1}^{n}(\Delta i)^2 \Rightarrow \boxed{\Delta i = y_i - y_{pred}}$$

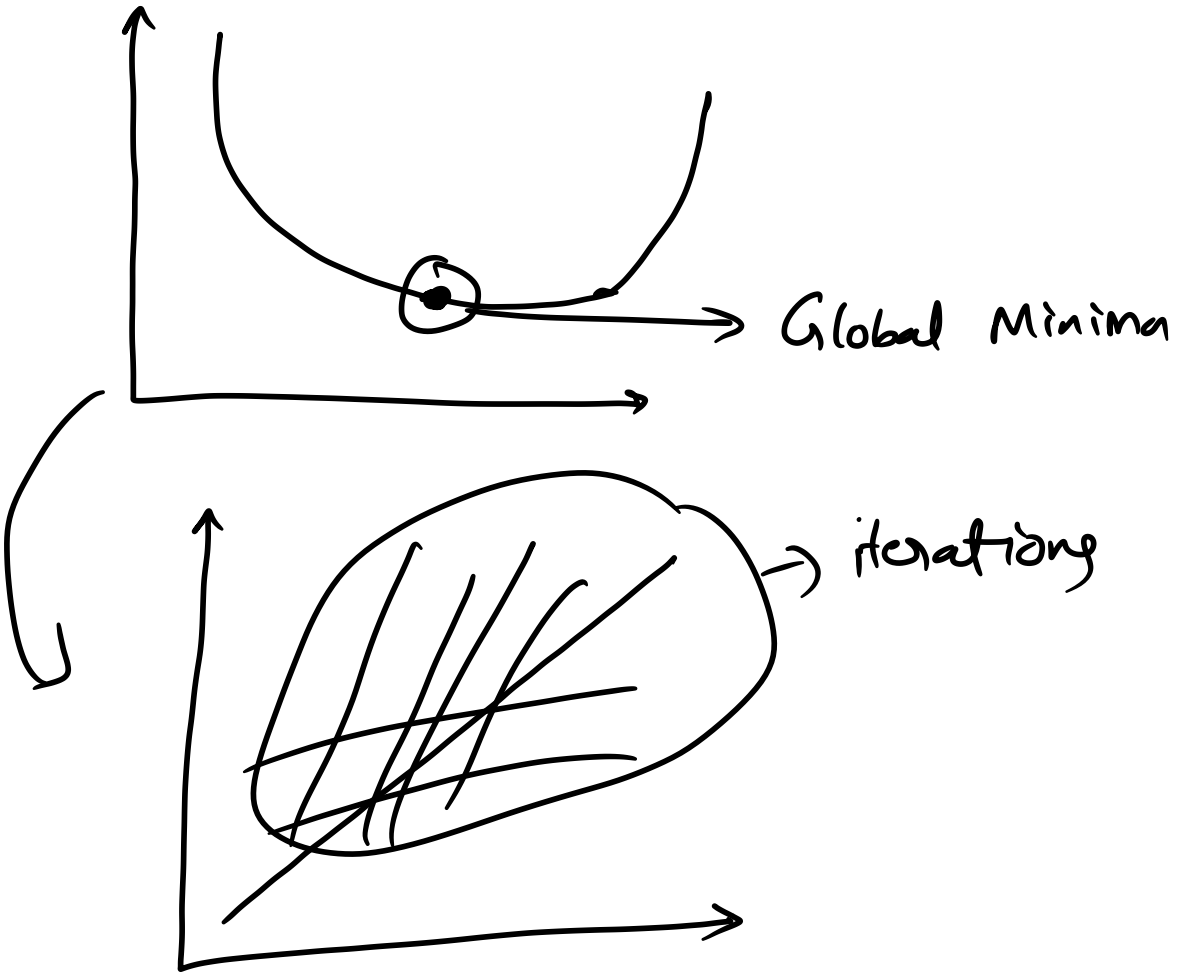
$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{pred})^2$$

$$\boxed{y_{pred} = mx_i + c}$$

$$mse = \frac{1}{n}\sum_{i=1}^{n}(y_i - (mx_i + c))^2$$

Cost Function

Gradient Descent



$$\frac{d}{dx} x^3 \Rightarrow x^n$$

$$= 3x^2 \Rightarrow n x^{(n-1)}$$

$$\frac{d}{dx} = \boxed{3x^2 + 2y^2}$$

$$\frac{d}{dx} = (3 \times 2)x + 0$$

$$= 6x$$

$$\frac{d}{dy} = 4y$$

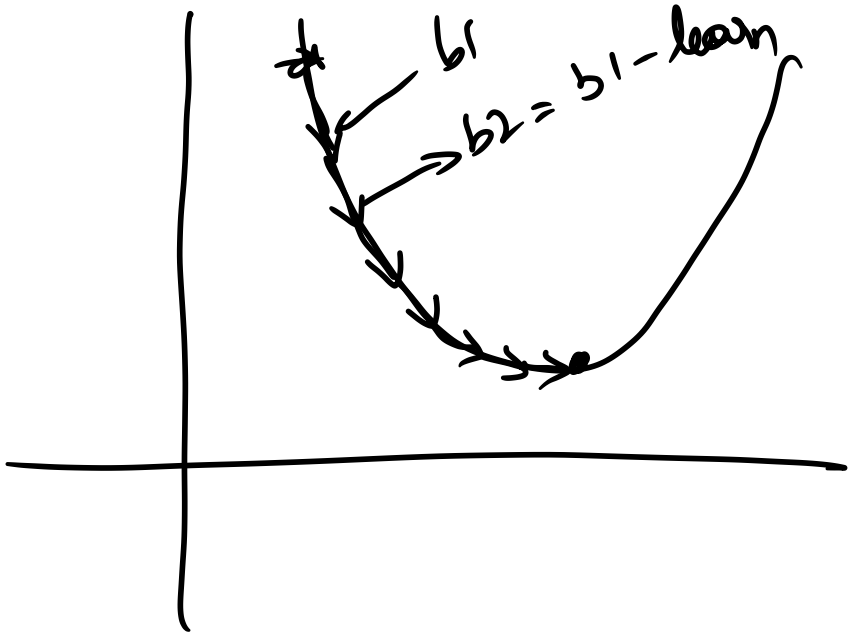
$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

$$\boxed{\frac{\partial}{\partial m} = \frac{2}{n} \sum_{i=1}^{n} -x_i (y_i - (mx_i + b))}$$

$$\frac{\partial}{\partial b} = \frac{2}{n} \sum_{i=1}^{n} -(y_i - (mx_i + b))$$

$$m = m - \text{learning rate} * \frac{\partial}{\partial m}$$

$$b = b - \text{learning rate} * \frac{\partial}{\partial b}$$



One Hot Encoding

town	area	price
Southern		
North		
East		
East		
South		
North		
West		

0
1
2
3
0

Categorical Variables

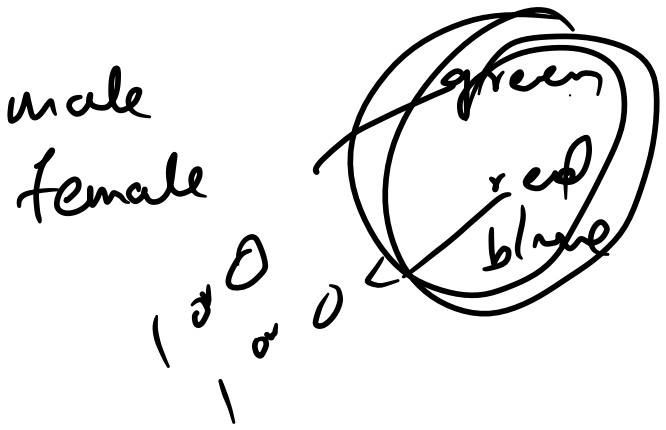
Nominal

Ordinal

kut, kphb, maisma

high → 2
medium → 1
low → 0

male
female



graduate → 0
master → 1
phd → 2

Linear Regression

1. Home prices
2. wheather
3. stock price

↳ predicted value is continuous

↓

Regression

Logistic Regression

↳ Email is spam or not

↳ will customer buy life insurance

↳ predicted value is categorical.

↓

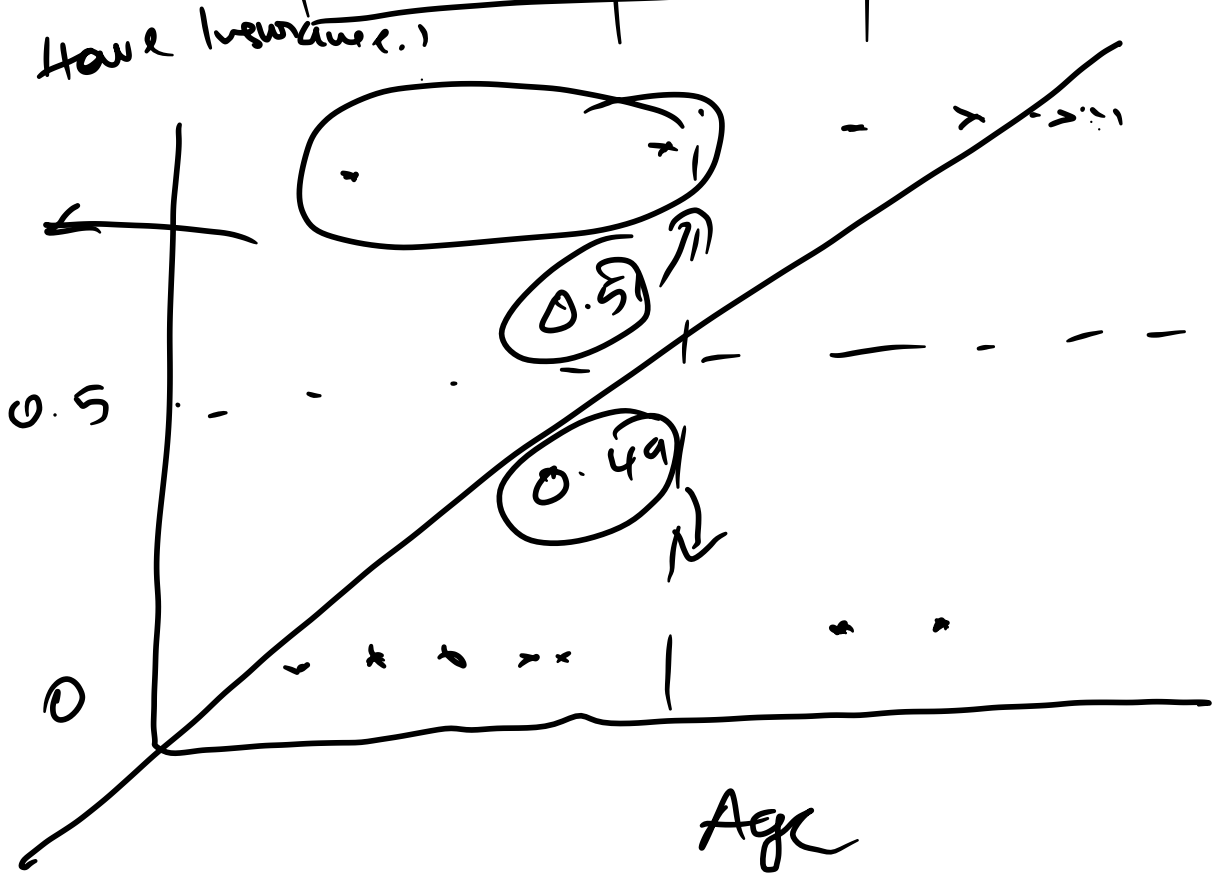
classification

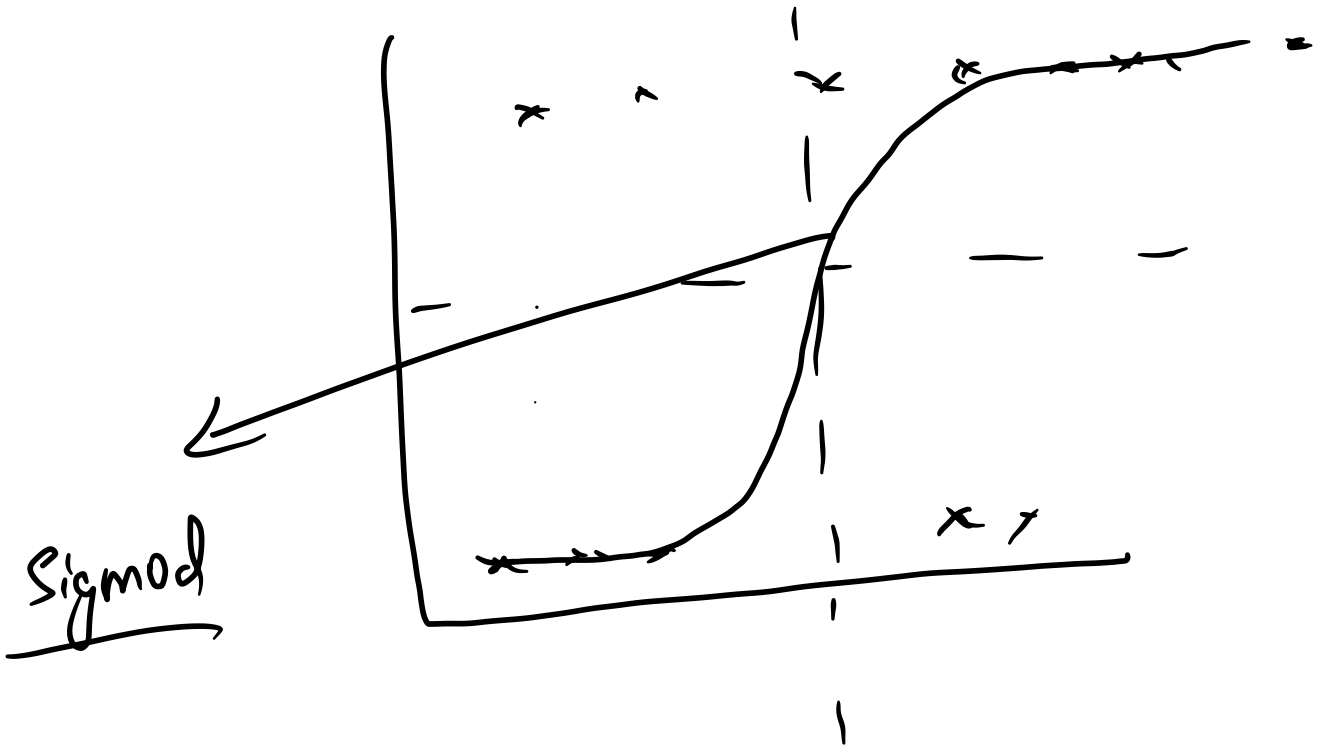
Logistic Regression

~~Binary~~ classification — Yes / NO

~~multiclass~~ classification — 1. Democratic, 2. Republican, 3. Independent

age	have_insurance
22	0
25	1
47	1
52	0
56	1
55	1
60	0
62	1

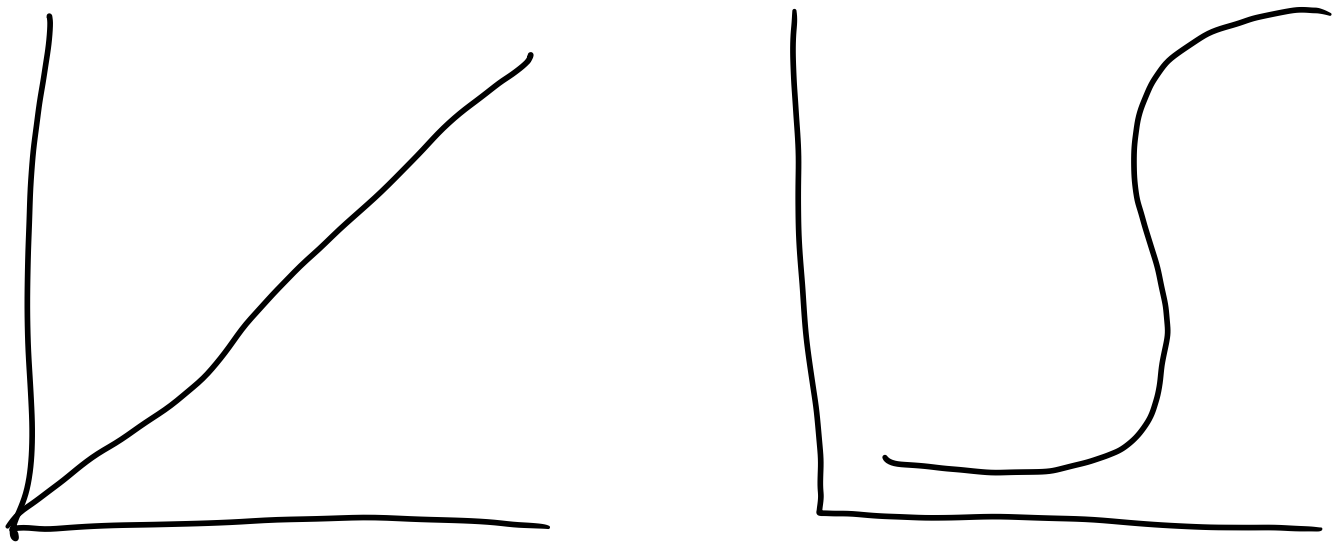




$$\text{sigmoid}(z) = \frac{1}{1 + e^{-z}}$$

e = Euler's number ~ 2.71828

Sigmoid function converts input into range 0 to 1



$y = mx + b$ $\quad z = \frac{1}{1+e^{-y}}$

$$y = \frac{1}{1 + e^{-(mx+b)}} \begin{cases} 0 \\ \text{or} \\ 1 \end{cases}$$

Logistic