Linear Regression (- supervised learning, - Regression)

1 - SL algos usally try to find the relationship between two vars (input and output) we try to find a function that maps one the other, Liner Regretion it trying to find a linear relationship between two variables (input and output)

2 ☆ we fit a linear equation to the data ⇒ 3 ☆ by miimizing the sum of squares of distances between data points (true val) and regretion line (estimate)

$$Y_i = \beta_0 + \beta_1 X_i$$

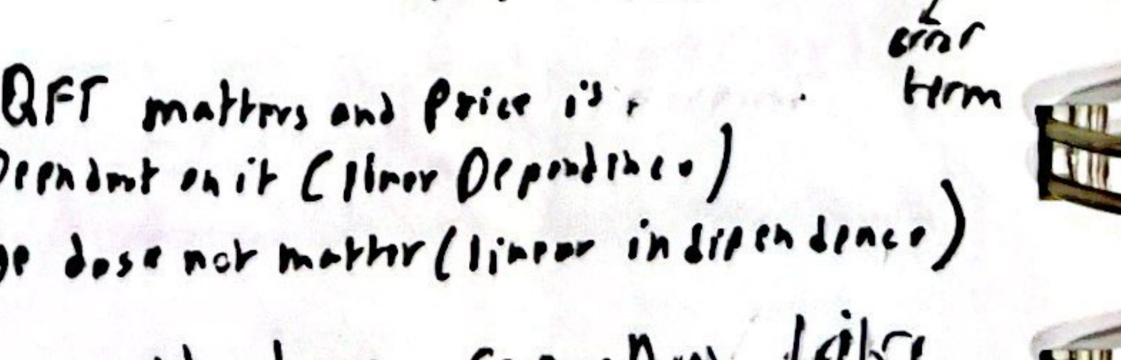
- Y_i ↙ Dependent variable (response var)
- β_0 ↓ const/ intercept
- β_1 ↘ slope/ coifficent
- X_i → indipendent variable (pridictor)

this simply minimize error

$$SSE = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

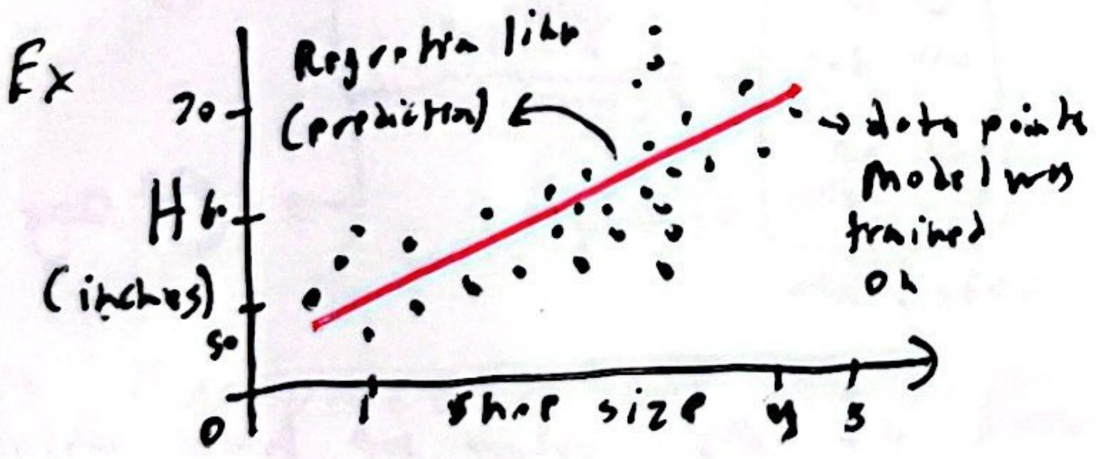
y_i = true val (datapoints)

$\bar{y}$ = prediction (line)



we want $SUM(y - \bar{y})^2 \rightarrow Min$

4 ☆ Simple ex of linear relashonship is height and shoe size or person where regretion fit can tells us for one unit of shoe size increase a person will be on avg 2 inches taller



5 ☆ you can make the model more complex and fit to multidimentional data for ex in shoe size EX we can add gender + Ennicity (X_1, X_2) to get better predictions

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon_0$$

error term

6 ☆ many ML algos Like NN are just Extentions of this concept

7 Ex: ☆ SQFT matters and Price is Dependent on it (linear Dependence) ☆ age dose not matter (linear independence)

7 ☆ in Math or Behind the scenes you would have something Like

Data

SQFT (x)	Price (y)
1000	200,000
1500	250,000
2000	300,000

SQFT vs Price of house

then we would find the line → Price = w · SQFT + b

w = slope (price increase per SQFT)

b = intercept (price at SQFT = 0)

8 ☆ then use formulas to find w, b and get: $y = 100x + 100,\ldots$

↳ 1 SQFT ⇒ price + = 100

↳ SQFT = 100,00 ← 19 time says