Linear Regression (- supervised learning, - Regression)

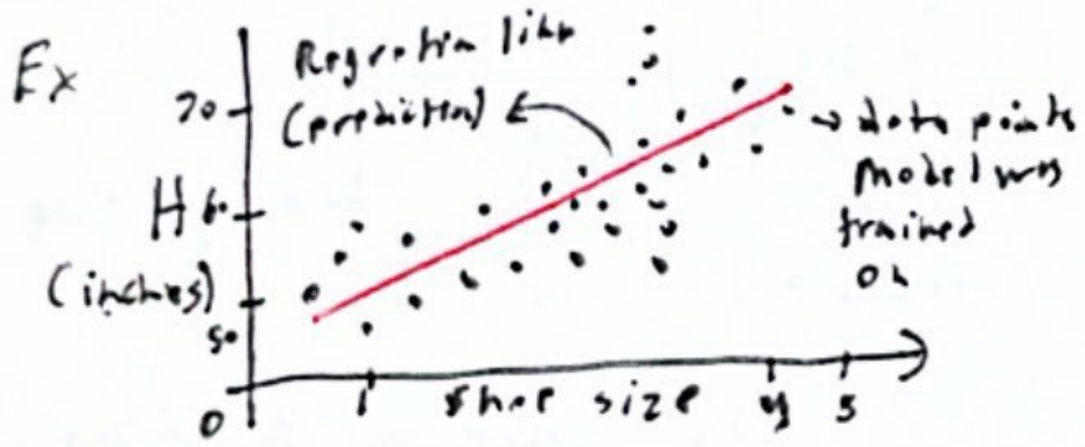
1 - SL algos usually try to find the relationship between two vars (input and output) we try to find a function that maps one the other, Liner Regretion it trying to find a linear relationship between two variables (input and output)

2 ☆ we fit a linear equation ⇒ 3 ☆ to the data

$$Y_i = B_0 + B_1 X_i$$

- Y_i: Dependent variable (response var)
- B_0: const/ intercept
- B_1: slope/ coifficent
- X_i: indipendent variable (pridictor)

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 increse a person will be on avg 2 inches taller



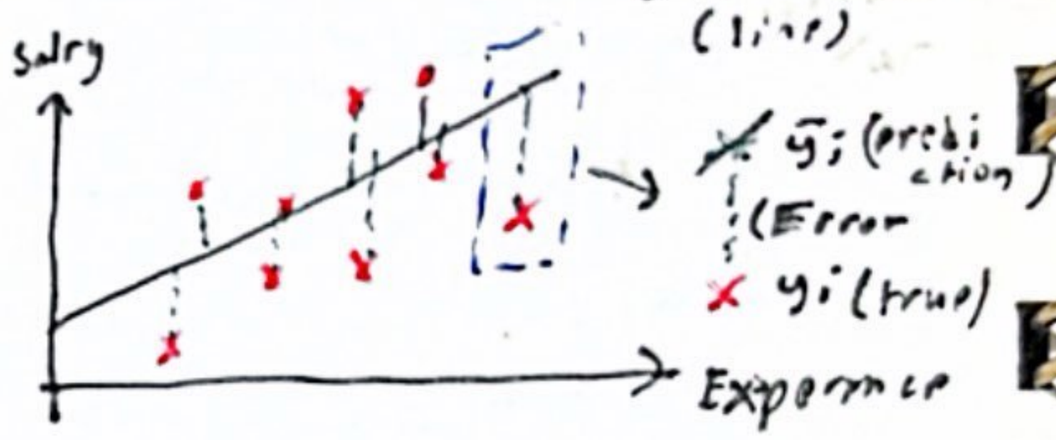
3 ☆ by minimizing the sum of squares of distances between data points (true val) and regretion line (estimate)

this simply minimize error

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

y_i = true val (data point)

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



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

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

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \cdots + \varepsilon_0$$

(ε_0: error term)

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

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

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

SQFT vs Price of house

Data

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

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

w = slope (price increase per SQFT)

b = intercept (price at SQFT = 0)

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

↳ +1 SQFT ⇒ price +: 100

↳ SQFT = 100,00 ← 19 ← line says

Scanned with CamScanner