

25/1/25

ML

Given a dataset for performing linear regression with one variable.

Example: Used car price prediction

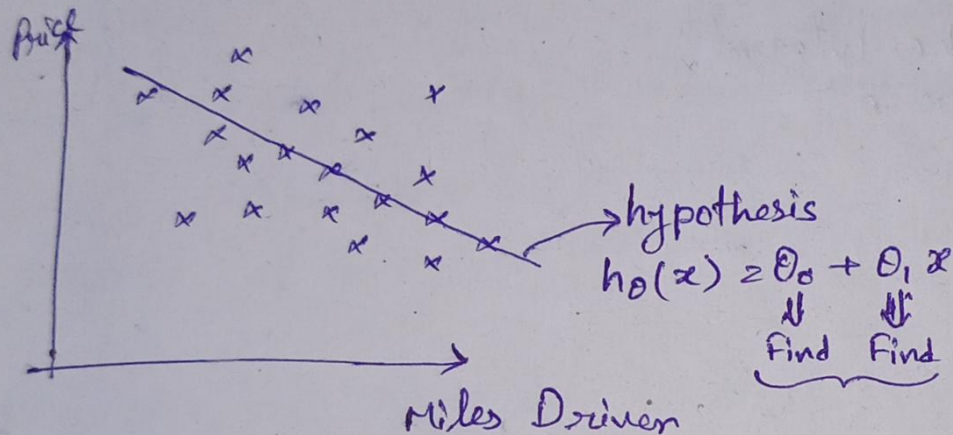
Features		o/p
Price		
I/p → Miles Driven	Engine Capacity	

But for simplicity purpose we are taking only one i/p.

Miles Driven	Price
1000	2.5
2520	3.6
⋮	⋮
⋮	⋮
⋮	⋮

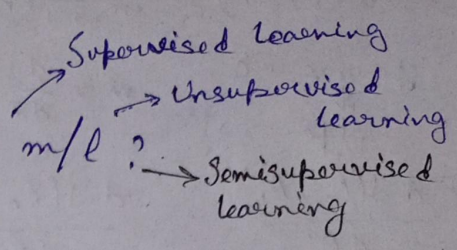
let, 2D matrix
(1000, 2.5)
(2520, 3.6)

$$y = mx + c$$
$$\geq 0, x + 0$$

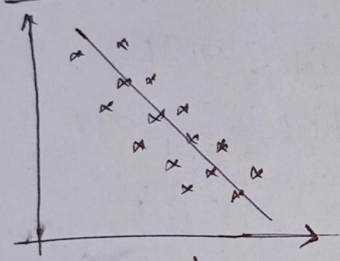


CSV (Comma Separate Vector) can store huge no. of datas in tabular form.

- 1) What is machine learning?
- 2) What is learning?
- 3) What are the different types of m/l?
- 4) What is Supervised Learning?
- 5) What type of problem we can solve using Supervised learning?
 - ↓ Regression
 - ↓ Classification
- 6) What is Regression problem explain with a real-life example?

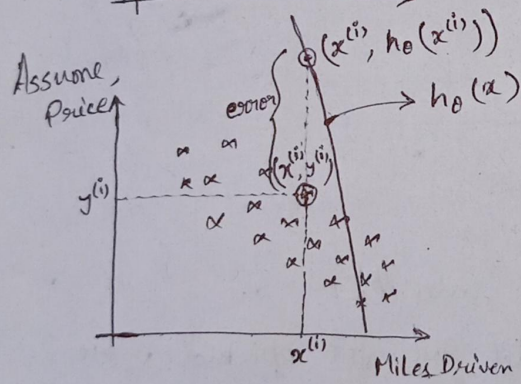


Linear Regression with one variable



$$h_0(x) = \theta_0 + \theta_1 x$$

Initially choose θ_0 and θ_1 randomly.



$$h_0(x) = \theta_0 + \theta_1 x$$

$$\text{error} = (h_0(x^{(i)}) - y^{(i)})^2$$

~~We only need +ve values~~

The distance can be -ve but we only need magnitude, but we can't do mod(1). Because, for later we have to do derivative but mod ~~make~~ can make problem at then. So, we can do square.

Now, assume there are m data samples.

$$\text{So, total no. of errors} = \frac{1}{m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)})^2$$

loss function
mean squared error (MSE)

Minimize Error

$$\text{Minimize } J(\theta_0, \theta_1)$$
$$\theta_0, \theta_1$$

$$J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)})^2$$

$$y = mx + c$$

↓
slope

$$m = \frac{dy}{dx}$$

$$\frac{dJ(\theta_0, \theta_1)}{d\theta_0}$$

$$\frac{dJ(\theta_0, \theta_1)}{d\theta_1}$$

Updating the value of θ_0, θ_1 :

$$\theta_0 = \theta_0 - \alpha \left(\frac{dJ(\theta_0, \theta_1)}{d\theta_0} \right) \rightarrow \text{gradient}$$

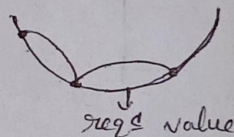
$$\theta_1 = \theta_1 - \alpha \cdot \frac{dJ(\theta_0, \theta_1)}{d\theta_1}$$

α = Learning rate

If $\alpha \geq 1$ then it means α is too high. That means it is seeing the slope and making a high jump. It can make problem.

Steps:

- ① (i) Take Random points
- (ii) Calculate errors
- (iii) Calculate gradients
- (iv) Calculate α -value
- (v) Calculate new value for θ_0 and θ_1
- (vi) Repeat from step (i) until we get optimal value



Performing Partial Derivative:

For θ_0

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)})^2$$

$$h_0(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

$$\frac{d}{d\theta_0} J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m \frac{d}{d\theta_0} (\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2$$

$$= \frac{1}{2m} \sum_{i=1}^m \frac{d(\theta_0 + \theta_1 x^{(i)} - y^{(i)})^2}{d(\theta_0 + \theta_1 x^{(i)} - y^{(i)})} \times \frac{d(\theta_0 + \theta_1 x^{(i)} - y^{(i)})}{d\theta_0}$$

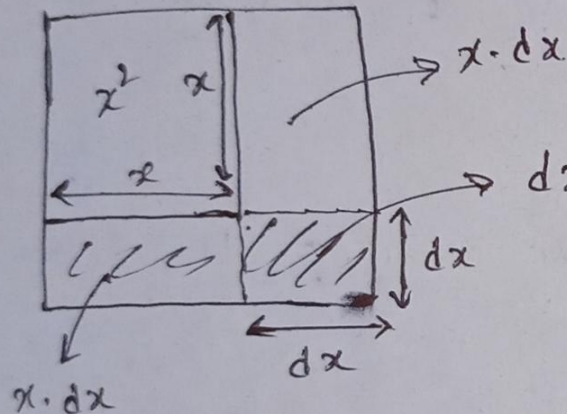
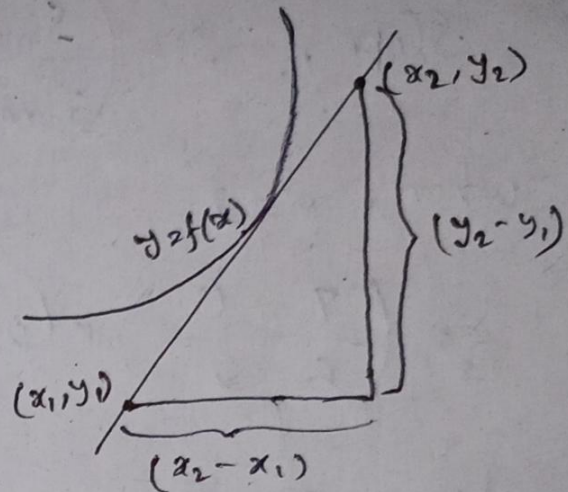
$$= \frac{1}{2m} \sum_{i=1}^m 2(\theta_0 + \theta_1 x^{(i)} - y^{(i)}) (1 + 0 - 0)$$

$$= \frac{1}{m} \sum_{i=1}^m (\theta_0 + \theta_1 x^{(i)} - y^{(i)})$$

$$\left[\frac{d}{d\theta_0} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)}) \right]$$

Now, for θ_1

$$\frac{d}{d\theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$



$$\begin{aligned} (x+dx)^2 &= x^2 + x dx + x dx + dx^2 \\ &\Rightarrow (x+dx)^2 - x^2 = 2x dx \end{aligned}$$

very small
-value

$$\Rightarrow \frac{(x+dx)^2 - x^2}{dx} = 2x$$

$$\begin{aligned} \tan \theta &= \frac{\text{height}}{\text{base}} \\ &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{dy}{dx} \end{aligned}$$

ML

30/1/2025

Total Error $\rightarrow \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$

Avg. err \downarrow

Mean Square Error $\rightarrow \boxed{\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2} \rightarrow \text{Loss function}$

⊗ Gradient Descent Algorithm

⊞ Linear Regression with multiple features:

Input			output
Miles Driven (x_1)	Engine Capacity (x_2)	Fuel Type (x_3)	Price (y)

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Fix
 $x_0 = 1$

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$
$$= \theta^T x$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n+1}$$

$$x = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

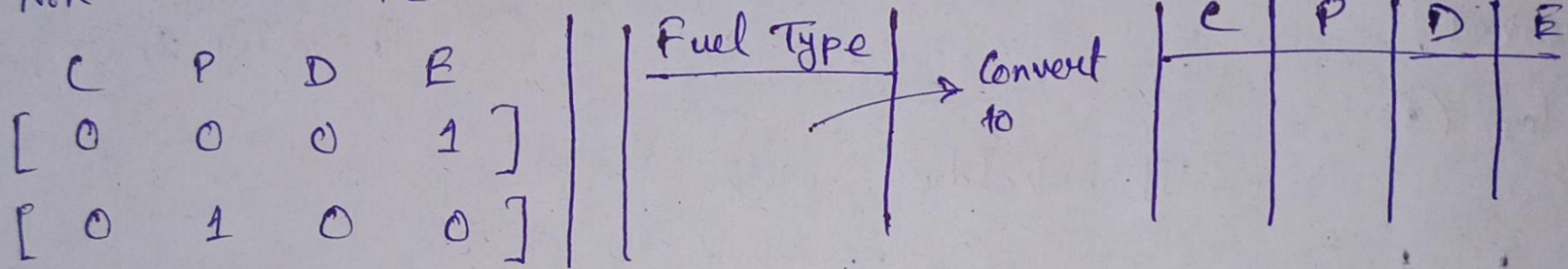
$$\begin{bmatrix} \theta_0 & \theta_1 & \dots & \theta_n \end{bmatrix}_{(1 \times n)} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}_{(n \times 1)}$$

Repeat Control Convergent :

$$\theta_j = \theta_j - \alpha \frac{d}{d\theta_j} J(\theta) \quad J = 0 \text{ to } n$$

ONE HOT ENCODING :

• Non-numerical Feature $\xrightarrow{\text{to}}$ Vector



Normalization of Features : / Feature Scaling

[0 - 1]

Age	Salary
31	80000
32	80001
40	240000

min (age)
max (age)

$$\frac{\text{age} - \min}{\max - \min} \rightarrow [0 - 1]$$

ML

ML

Miles Driven	Engine Capacity	Year	No. of Pre owners	Fuel Types				
2000				CNG	C	D	P	B
	[-1 to 1]	[-1 to 1]	[-1 to 1]	Des.	0	1	0	0
				Pet.	0	1	0	0
				Bat.	1	0	0	0

$\square \leq 0.5 \rightarrow \text{Class 0}$

$\square > 0.5 \rightarrow \text{Class 1}$

0.5 is threshold value. If the probability value is 0.5 then the machine will think that there has 50% chances for both.

Qn1 What is the role of the sigmoid / logistic fn to determine the class?

• Cost fn and Loss fn are same.

Regularization: Overfitting Problem

Under fit, Perfect fit, Overfit

ML

8/2/25

Classification using Decision-Tree

Features

Age, Income, Student? Credit Rating

The prediction is basically ~~buy~~ computer or not.

Example : Age Income Student Credit Rating buy By Comp.

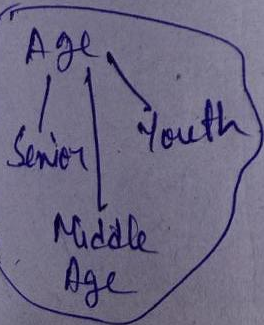
Youth

High

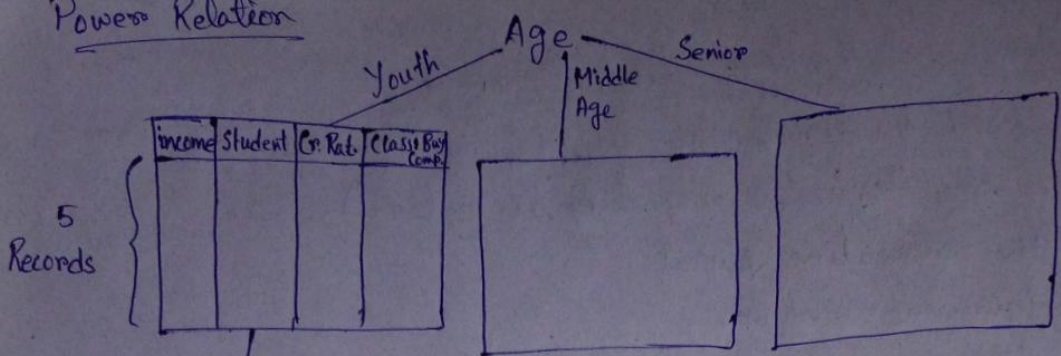
No

Fair

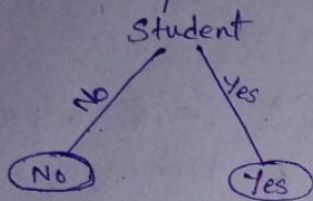
No



Power Relation



Biased data So, again splitting



① if age = 'middle age' then Buy PC = Yes

② if age = 'youth' and student = 'no' Buy PC = No

If ~~no~~ after splitting all ~~that~~ attributes, there left no attribute then go for majority voting. If the majority is 'yes' then all will be 'yes' and vice-versa.

Attribute Selection for Splitting

Information gain / Entropy

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i)$$

$\therefore m = \text{no. of classes}$

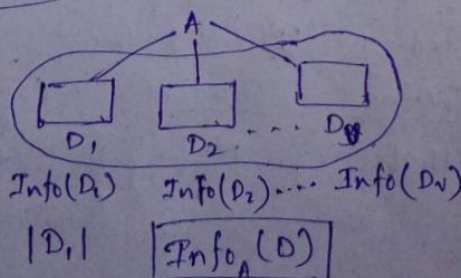
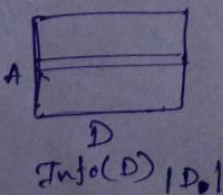
For our case,

$m = 2$ → Buy computer = Yes & No
 p_1 → Yes
 p_2 → No

• The unit of entropy is bits.

$$IC \propto \frac{1}{P(A)}$$

Information Content



$$\text{Info}_A(D) = \frac{|D_1|}{|D|} \text{Info}(D_1) + \frac{|D_2|}{|D|} \text{Info}(D_2) + \dots + \frac{|D_n|}{|D|} \text{Info}(D_n)$$

$$= \sum_{j=1}^n \frac{|D_j|}{|D|} \text{Info}(D_j)$$

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D)$$

$$\text{Gain}(\text{Age}) = \text{Info}(D) - \text{Info}_{\text{Age}}(D)$$

$$\text{Info}(D) = - \sum_{i=1}^n (p_i \times \log_2(p_i))$$

$$= -p_1 \log_2(p_1) - p_2 \log_2(p_2)$$

$$= -\frac{9}{14} \log_2\left(\frac{9}{14}\right) - \frac{5}{14} \log_2\left(\frac{5}{14}\right) = 0.9700$$

$$= -0.64 \log_2(0.64) - 0.36 \log_2(0.36)$$

$$= (-0.64) \times (-0.15) - (0.36) \times (-0.47)$$

$$= 0.288 + 0.3612 = 0.288 + 0.3612$$

$$= 0.28$$

$$\text{Info}_{\text{Age}}(D) = \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] +$$

$$\frac{5}{14} \left[-\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) \right] +$$

$$\frac{4}{14} \left[0 - \frac{4}{4} \log_2\left(\frac{4}{4}\right) \right]$$

$$= \frac{10}{14} (0.97 + 0.97)$$

$$= 0.69285$$

$$\text{Gain}(\text{Age}) = 0.97 - 0.69 = 0.2465$$

$$\text{Info}_{\text{Age}}(D) = \sum_{j=1}^3 \frac{|D_j|}{|D|} \text{Info}(D_j)$$

$$= \sum_{j=1}^3 \frac{|D_j|}{|D|} \text{Info}(D_j)$$

$$= \frac{|D_1|}{|D|} \text{Info}(D_1) + \frac{|D_2|}{|D|} \text{Info}(D_2) + \frac{|D_3|}{|D|} \text{Info}(D_3)$$

$$= \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right] + \frac{4}{14} \left[-\frac{4}{4} \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \log_2\left(\frac{0}{4}\right) \right]$$

$$+ \frac{5}{14} \left[-\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) \right]$$

$$= 0.6928$$

$$\text{Gain}(\text{Age}) = 0.97 - 0.69 = 0.28$$

Gain of Income ≈ 0.029 bits

Gain of Student ≈ 0.151 bits

Credit Rating ≈ 0.048 bits