Linear Regression

Agenda

1. Linear Regression
2. Cars24 Case
3. Intuition lin.reg
4. Maths → Algebraic
5. Sklearn → Code

Cars 24 Problem Overview

Data scientist at **Cars 24** → Sells pre-owned cars

Resale Price

→ To automate pricing the old car

Given car features

info.

(Make, Model Mileage , Odometer, Service History etc.) →

Predictors / Features

ML Model

Price of car

input var.
independent

output / dependent target

$y \in R$

What do you think about the nature of Car Resale price prediction?

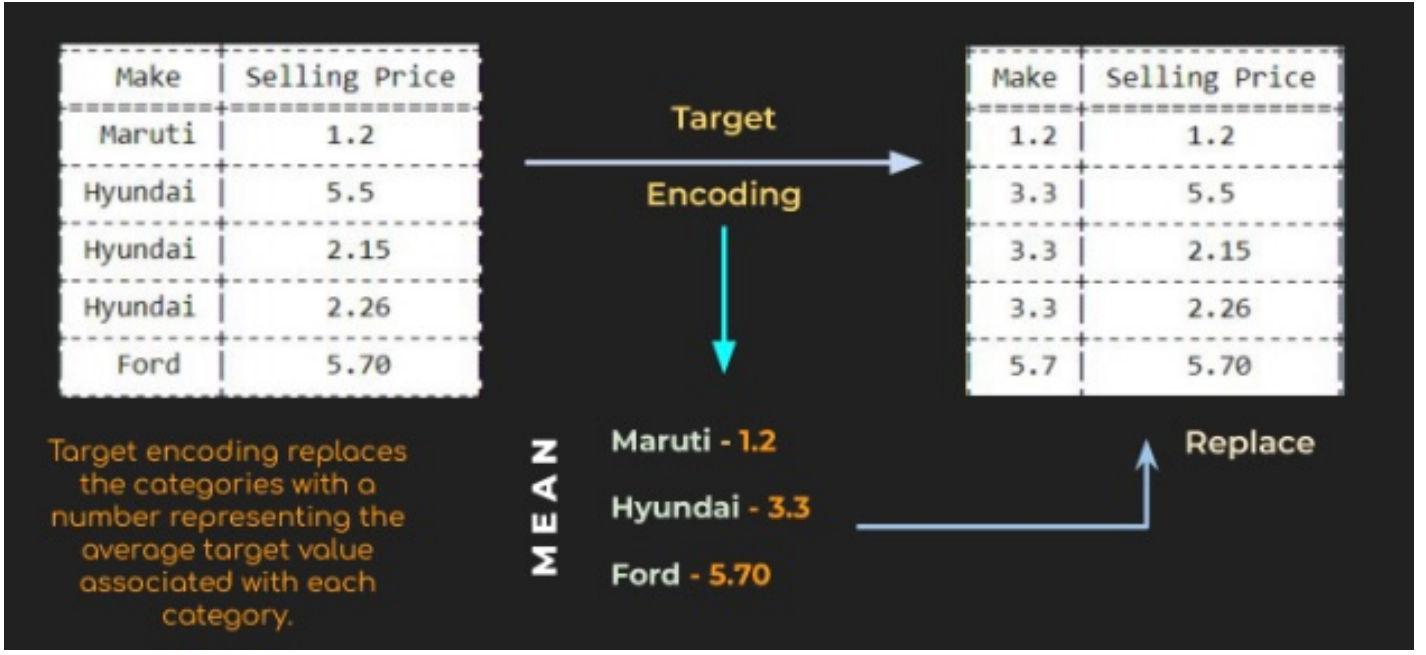
110 users have participated

"Supervised"

A	Regression	85%
B	Classification	9%
C	Clustering	5%

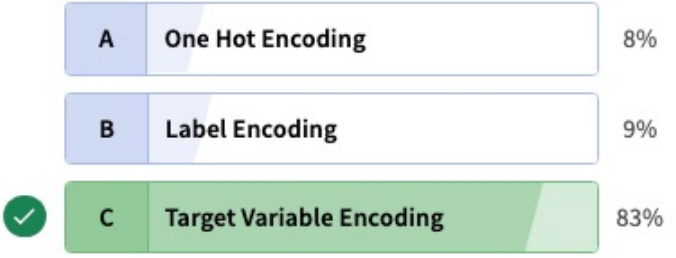
OHE → 3300 New cols/features ⇒ Target Encoding

"Curse of Dim^n"

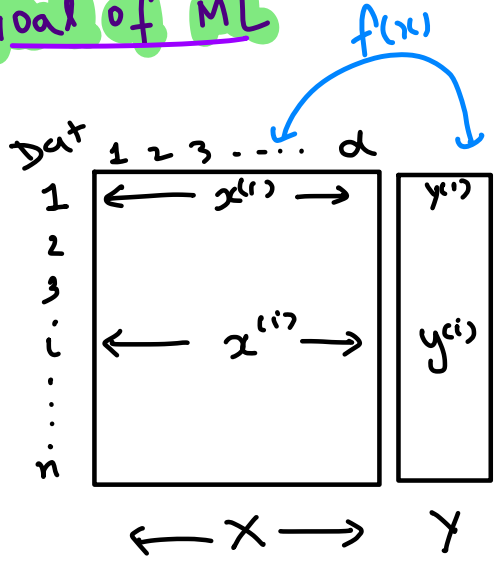


How do you think we should handle the large number of categories in make and model column?

117 users have participated



Goal of ML



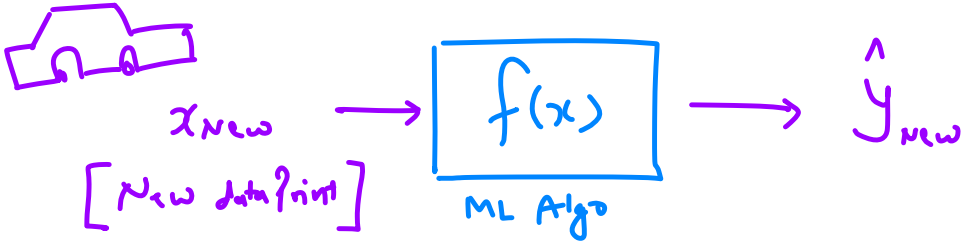
$y^{(i)}$ → actual s.p

$x^{(i)}$ → ML Algo / ML Model → $\hat{y}^{(i)}$ [Predicted s.p]

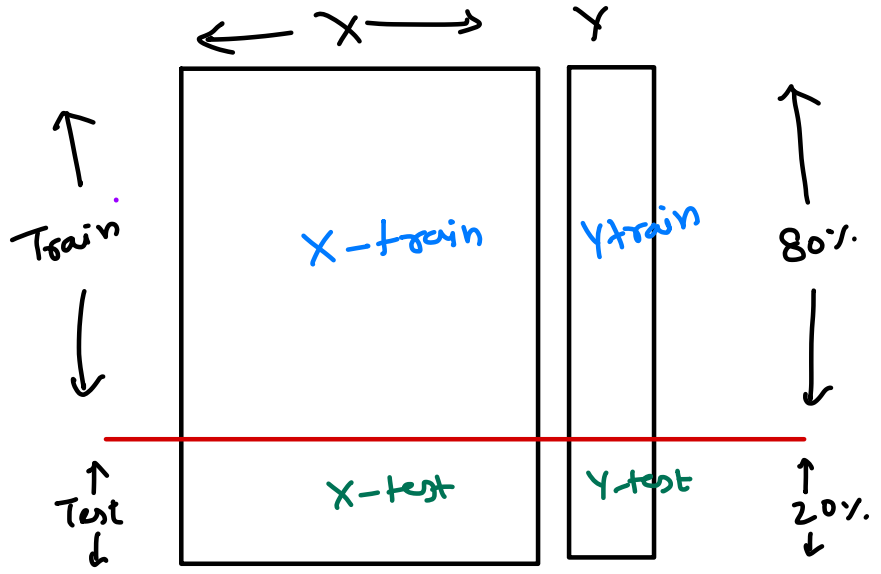
Ideally,

$$y^{(i)} \approx \hat{y}^{(i)}$$

original predicted.



Train Test Split



ML Train

↓

(X-train , y-train)

$f(x)$ / $h(x)$

"Evaluation"

X-test → $h(x)$ → $\hat{y}_{pred}$

↕ Compare

Y-test

Intuition L.R

→ Univariate Lin. Reg [1 input var]

→ Multivariate Lin. Reg [> 1 input var]

input → engine ⟶ output → Price

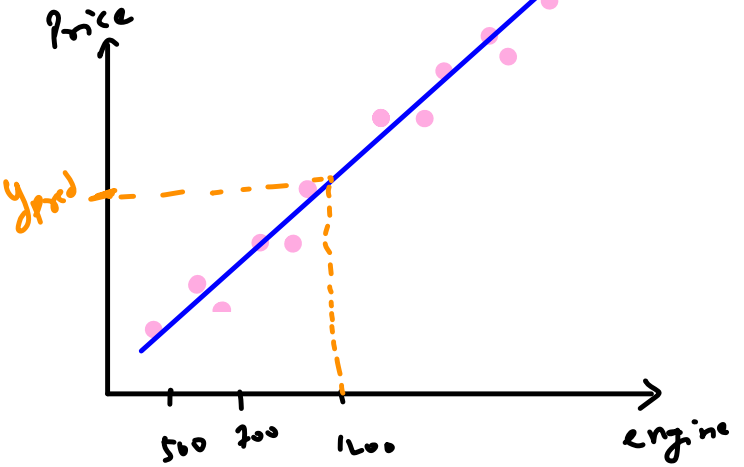
NewGr (1200 cc)

↓

input this in line eq.

↓

get Price



$x^{(i)}_{new}$ ⟶ (line eq $h(x)$) ⟶ $y^{(i)}_{new}$ (predicted)

St. line $y = mx + c$

$$y = w_1 \cdot x + w_0$$

w_1: weight, x: engine, w_0: intercept/bias

For. eg (Assume)

$$\hat{y} = \underbrace{200}_{w_1} \cdot x + \underbrace{50000}_{w_0}$$

(1200 cc)

Price $= 200 \times 1200 + 50K$

$= 2.9L$

2-input features

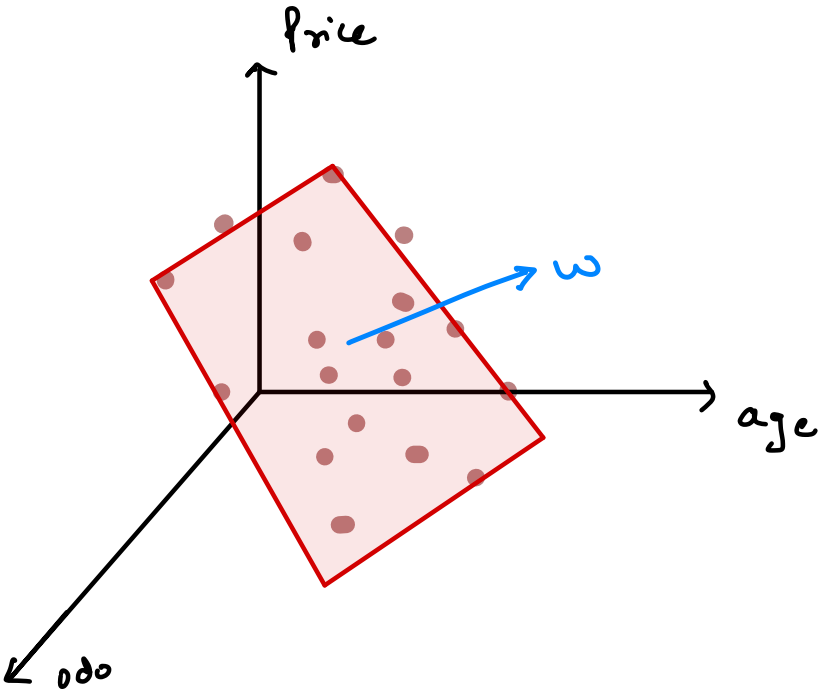
age , odometer

x_1	x_2	
age	odo	y

$$y = w_1 x_1 + w_2 x_2 + w_0$$

age, odo

weight, weight, intercept



$$y = -10000 x_1 - 10 \cdot x_2 + 5000$$

d- features

$$y = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + w_0$$

$$\hat{y} = w^T x + w_0$$

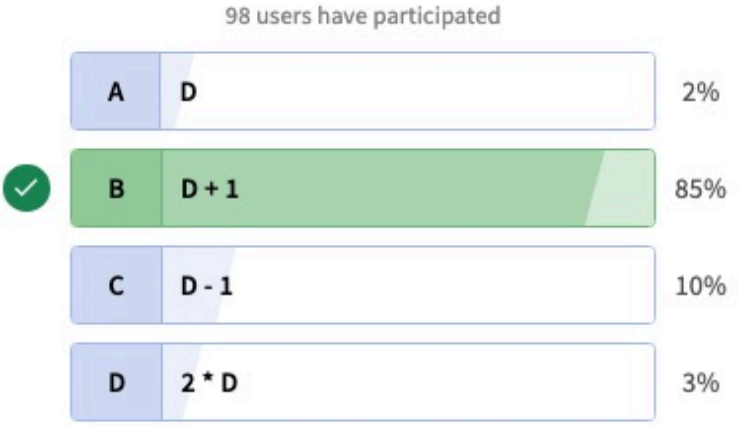
$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

1 feature $\longrightarrow$ 2D line

2 features $\longrightarrow$ 3D plane

d feature $\longrightarrow$ hyperplane (d+1)

If your data contains d features, how many dimensions will be required to fit the hyperplane through that data?



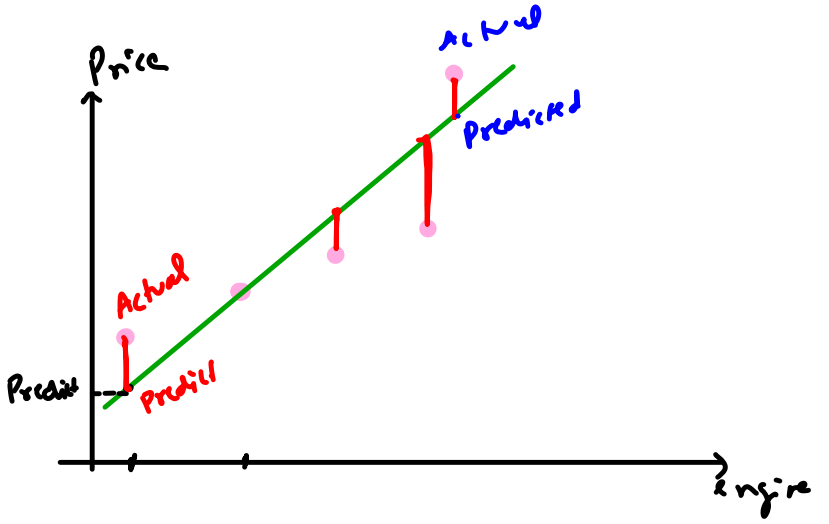
Evaluation Metric

$x^{(1)} \rightarrow y^{(1)} - \hat{y}^{(1)} = e^{(1)}$

$x^{(2)} \rightarrow y^{(2)} - \hat{y}^{(2)} = e^{(2)}$

$\vdots$

$x^{(5)} \rightarrow y^{(5)} - \hat{y}^{(5)} = e^{(5)}$



$$\text{Total error} = \frac{1}{n}\sum_{i=1}^{n} e^{(i)}$$

$$\text{Total error} = e^{(1)} + e^{(2)} + \cdots + e^{(5)}$$

$$= 3 + 0 + (-2) + (-4) + (3)$$

$$= 0$$

$|3| = 3$

$|0| = 0$

$|-2| = 2$

$|-4| = 4$

$|+3| = 3$

Error = $|2|$

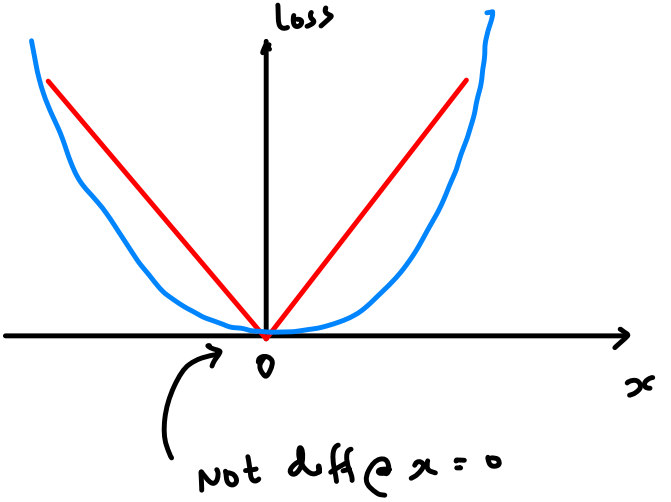
$\Rightarrow$

$$\text{Error} = \sum_{i=1}^{n} |e^{(i)}| = \sum_{i=1}^{n} |y^{(i)} - \hat{y}^{(i)}|$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y^{(i)} - \hat{y}^{(i)}|$$

mean Sq. Error

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(y^{(i)} - \hat{y}^{(i)} \right)^2$$



MSE

M_1 9.62 ✓ Better

M_2 15.41

R2 Score → Next

In linear regression, if the MSE value is 0, it indicates:

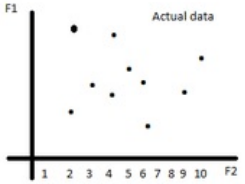
85 users have participated

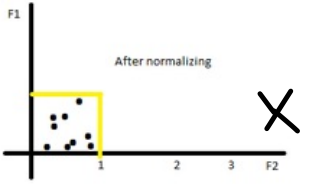
✓	A	The predicted values perfectly match the actual values.	69%
	B	The model has no predictive power and fails to explain the dependent variable.	15%
✗	C	The model has high bias and underfits the data.	8%
✗	D	The model has high variance and overfits the data.	7%

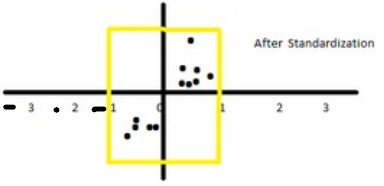
End Quiz Now

MinMaxScaling → images [0-255] [0-1]

StandardScaling → $\frac{x-\mu}{\sigma}$







$\mu = 0$
$\sigma = 1$

. transform

. inverse_transform

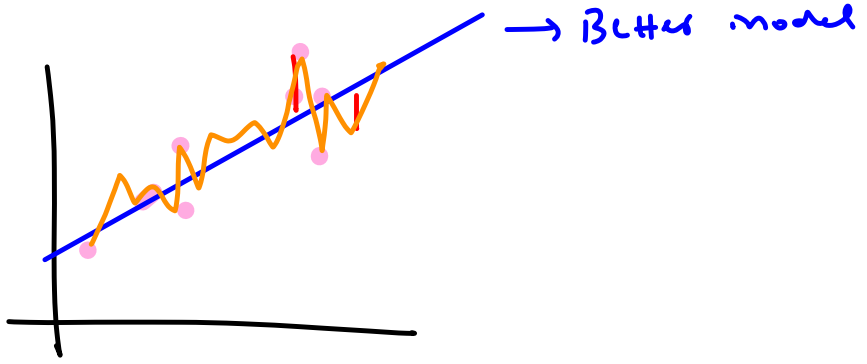
$$x' = \frac{x-\mu}{\sigma}$$

$$[x = x' \cdot \sigma + \mu]$$

→ 1.4 ↓ 8 yrs

age → -0.8 ↓ 4yrs

→ -2.1 ↓ 1yr.



$MSE \in (0, \infty)$

RMSE

$y = \omega x + \omega_0$

$(0.1) \cdot x + (0.1)$

$n = 1M$

$900K$ 100

$970K$ $50K$

100 d

2^{100} 2^d

$0 + 1 + 1 + 1$

2

4

8