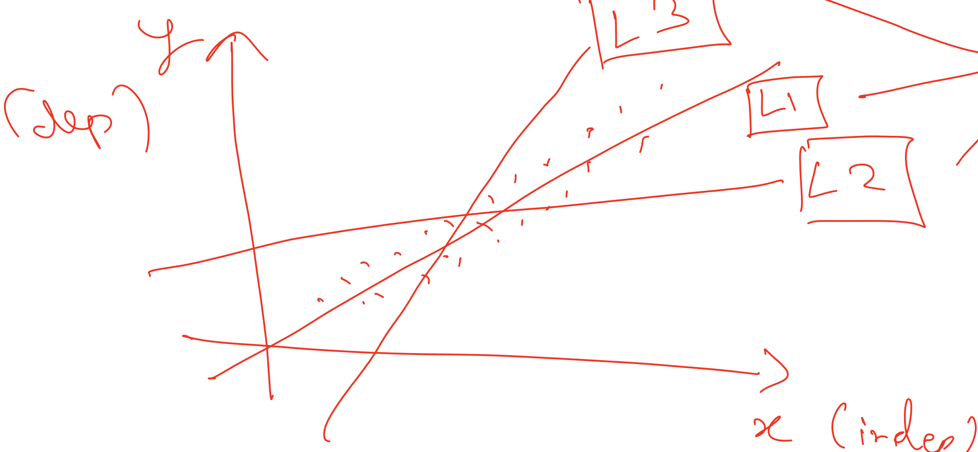
Last Class - 25 May

- → Linear Regression
- → Single variable (Univariate) Prediction
- → Gradient Descent (Batch GD)
- → Mean Squared Error
- → Derivatives & Gradients.
- → Cars 24 Price Prediction Problem using single variable

Today's class — May 27

1) Goodness of the Model (being trained or fit)
   → R2 (Coefficient of Determination)
2) Univariate (single variable) analysis – review
   → features
3) Linear Regression Library from Python
4) Multivariate Regression (Multiple Variables)
   ↓ features
5) Adjusted R2 →
6) Model Interpretability & Feature Importance

R2 Score



which is the best model ?

$y = mx + c$

$y = w_0 + w_1 x$

parameters

$\boxed{R2} = 1 - \dfrac{\text{SUM of residuals (SR)}}{\text{Total Sum of Squares (TS)}}$

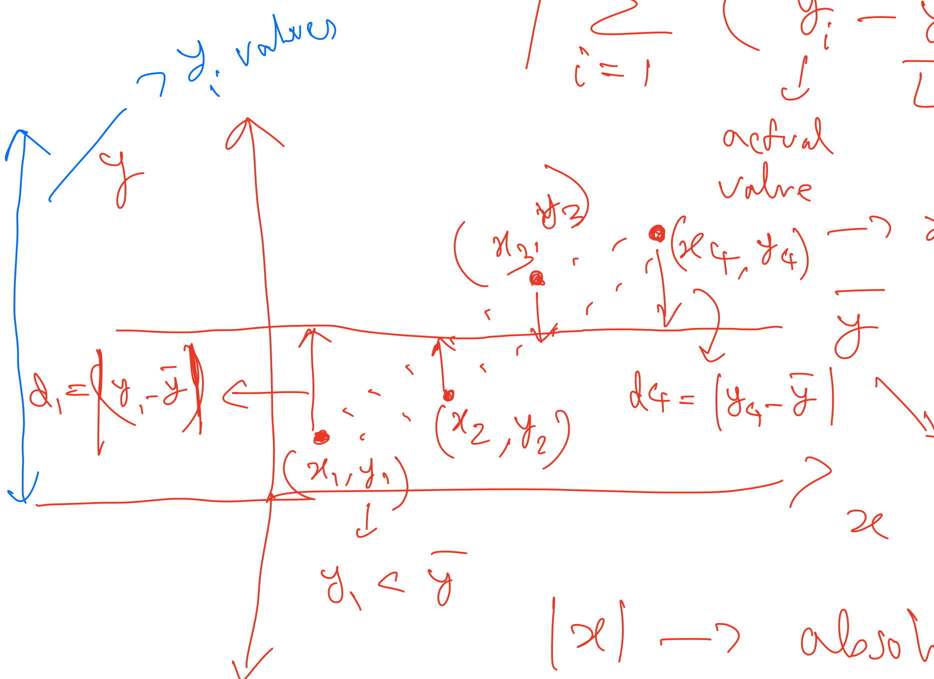
→ num = SR = $\sum_{i=1}^{N} (\hat{y}_i - y_i)^2$

$\hat{y}_i$ → pred, y_i → actual

→ den = TS = $\sum_{i=1}^{N} (y_i - \bar{y}_i)^2$ ✓

y_i → actual value

$\bar{y}_i$ → average of actual y_i's



→ $y_4 > \bar{y}$

$\bar{y}$ (y_dash)

average $\hat{y}$ (y_hat)

$|x|$ → absolute value of x

$x^2 \rightarrow +ve$ $d_1 \rightarrow y_1 - \bar{y} < 0$

$d_4 \rightarrow y_4 - \bar{y} > 0$

$$R2 = 1 - \frac{\overbrace{\sum_{i=1}^{N}}^{\text{sum}} (\hat{y}_i - y_i)^2}{\underbrace{\sum_{i=1}^{N} (y_i - \bar{y})^2}_{\text{sum}}} \quad \begin{matrix} \rightarrow \text{num} \\ \rightarrow \boxed{\text{den}} \end{matrix}$$

Best Scenario

when all prediction $(\hat{y}_i)$ is equal to ground truth / actual value (y_i), i.e $\hat{y}_i = y_i$ for every i

$$\hat{y}_i - y_i = 0 \quad \text{for every } i$$

$\rightarrow$ num $= 0$

$$R2 = 1 - \frac{0}{\text{something}} = 1$$

Worst Case Scenario

Say num $>$ den, then $\frac{\text{num}}{\text{den}} > 1$
$\hookrightarrow = 2$

then $R2 = 1 - \boxed{\frac{\text{num}}{\text{den}}} < 0$
$\hookrightarrow > 1$

$R2 = 1 - 2 = -1$

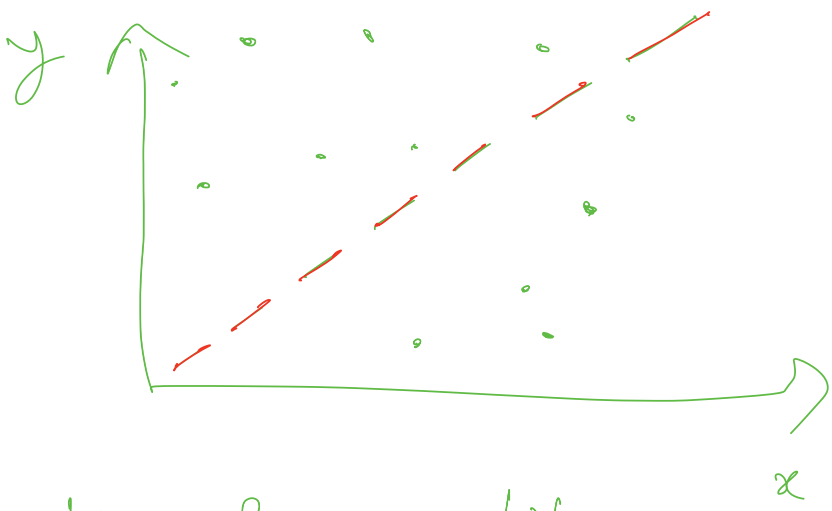
$\frac{\text{num}}{\text{dem}} \longrightarrow \infty$, which $\rightarrow R2 \longrightarrow -\infty$

Middle Case Scenario

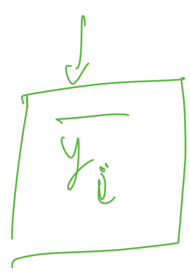
If num < den , $\frac{num}{den} < 1$ → 0.5

$R2 = 1 - 0.5$
$= 0.5$

$\Rightarrow R2 = 1 - \frac{num}{den} \rightarrow 0 < R2 < 1$



$\hat{y}_i$



Linear Regression Library :

coeff_ → w1

intercept_ → w0

X → (n_samples, n_features)

→ (19820 , 1)

Y → (n_samples, 1) → (19820, 1)

$x_i \rightarrow \quad y_i = \boxed{w_0 + w_1 x_i}$

w0, w1

100 samples $\longrightarrow$ 50 , 50

$\downarrow$ class A, $\downarrow$ class B

Random prediction $\longrightarrow$ 50% correct

✗ ML model1 $\longrightarrow$ 40% accuracy
$\downarrow$
bad job

✓ ML model2 $\longrightarrow$ 70% accuracy
$\downarrow$ might be useful
$\downarrow$ better than random

Outliers cause denom (Total Sum) in R2 score to be as large as possible, And hamper the predictions as well.

$$X = (x_1, x_2, \ldots, x_{17})$$

$$(1, x_1, x_2, \ldots, x_{17})$$

$$\downarrow \quad \downarrow \quad \downarrow \quad \quad \downarrow$$

$$\boxed{w_0} \quad w_1 \quad w_2 \quad \ldots \quad w_{17}$$

$$f(x_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots w_n x_{in}$$

$\boxed{h(x)}$

hypothesis function

np.dot(weight, ~~data~~ X)

$$\begin{array}{l} x_1 \\ x_2 \\ x_3 \end{array} \begin{bmatrix} x_{11} & x_{12} & x_{13} \cdots x_{1,18} \end{bmatrix} \quad \begin{bmatrix} w_1 & w_2 & w_3 & \cdots & w_{18} \end{bmatrix} \rightarrow w_1 \cdot x_{11} + w_2 x_{12} + \cdots + w_{18} \cdot x_{18}$$

$N \times 18$ $\quad$ 18×1

np.dot(X, weight)

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} \quad N \times 1$$

$$w_0 \cdot 1 + w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots w_d \cdot x_d$$

$[w_0, \; w_1]$

$$L(w_0, w_1) = (y - (w_0 + w_1 x))^2$$

(x, y)

$$L(w_0, w_1) = (y_i - (w_0 + w_1 x_i))^2$$

(x_i, y_i)

$$\left.\frac{\partial L}{\partial W_1}\right|_{x = x_i} = -2(y - \hat{y}_i) \cdot x_i$$

$$\boxed{\frac{\partial L}{\partial w_0}} = \frac{1}{N} \sum_{i=1}^{N} -2(y - \hat{y}_i) \cdot x_i \rightarrow x_{i1}$$

x_i

$$\left.\frac{\partial L}{\partial w_0}\right|_{x = x_i} = -2(y - \hat{y}_i)$$

$$\frac{\partial L}{\partial w_0} = \frac{1}{N} \sum_{i=1}^{N} -2(y - \hat{y}_i) \cdot \rightarrow x_{i0}$$

1

$$\frac{\partial L}{\partial w_j}\Big|_{j=0,1,2,\ldots D} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - h(x_i)\right)(-x_{ij})$$

$f(x_i)$ for w_j, x_{ij}

w_0, w_1

$w_0, w_1, w_2, \ldots$

$x_0, x_1, x_2, \ldots$

$$\frac{\partial L}{\partial w_3} = \frac{1}{N}\sum_{i=1}^{N} 2\left(y_i - \hat{y}_i\right)(-x_{i3})$$

$h(x_i) = w_0 + w_1 \cdot x_{i1} + w_2 \cdot x_{i2} + \cdots w_d x_{id}$

gradient vector $\rightarrow \left[\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \ldots \frac{\partial L}{\partial w_d}\right]$

$(18 \times 1) \rightarrow$ vector

$$w_0 = w_0 - \eta \cdot \frac{\partial L}{\partial w_0}$$

$$w_j = w_j - \eta \cdot \frac{\partial L}{\partial w_j}$$

η → learning rate

R2 score

Adjusted R2 score

Adjusted R2 $\rightarrow \left[1 - \frac{(1-R^2)(m-1)}{m-d-1}\right]$

m → no of samples / data points

d → number of dim / features

Case 1: $\underline{d \text{ increases}}$, not significant change in R2

$d \uparrow$, R2 $\underline{\text{constant}}$ /

$d \uparrow$, $m-d-1 \Downarrow$ dec

$\frac{(1-R^2)(m-1)}{m-d-1 \downarrow} \uparrow$ inc

adj R2 $\downarrow$ dec

$$\text{adj } R2 = 1 - \left(\frac{(1-R^2)(m-1)}{m-d-1}\right) \downarrow$$

dec

Case 2: $\underline{d}$ increases, AND $\underline{R^2} \uparrow$ increases significantly

$d \uparrow$, $m-d-1 \downarrow \implies \frac{(1-R^2)(m-1)}{\cancel{(m-d-1)}} \uparrow$ ×

$\boxed{R2 \text{ increase} >> d \uparrow}$ × R2 $\uparrow$

$\boxed{0 < R2 < 1}$

R2 increase < d increase

$d \uparrow$, $m-d-1 \downarrow$

R2 $\uparrow$, $1-R2 \downarrow$ ↙ $\boxed{\frac{\downarrow(1-R^2)^{\uparrow}(m-1)}{(m-d-1)\downarrow}} \downarrow$

$$\text{adj } \underline{R2} = \left[1 - \frac{(1-R^2)(m-1)}{(m-d-1)}\right] \uparrow$$

Feature Importance

18 features

w_2 w_3

$$y = w_0 + w_1 x_1 + 2 x_2 + \underline{10 x_3} + w_4 x_4 + \dots w_{18} \cdot x_{18}$$

$w_2 = 2$, $\quad w_3 = 10$

~~$x = 3$~~

Let's say $x_2 \xrightarrow{\text{inc}} 1$ unit

$y \xrightarrow{\text{inc}} 2$ units

$$y = \boxed{k} + 2x_2$$

$\boxed{k} \rightarrow (w_0 + w_1 x_1 + w_3 x_3 + \dots w_{18} x_{18})$

$x_2 = 3$

$y = k + \cancel{2} \times 3$

$x_2 = 4$

$y = k + 2 \times 4$

$\rightarrow y \xrightarrow{\text{inc}} 2$ units

$x_3 \xrightarrow{\text{inc}} 1$ unit, $\quad y \xrightarrow{\text{inc}} 10$ units

$$y = w_0 + w_1 x_1 + (-5) \cdot x_2 + (-4) \cdot x_3 + w_4 \cdot x_4 + \dots w_{18} \cdot x_{18}$$

$x_2 \xrightarrow{\text{inc}} 1$ unit, $\quad y \xrightarrow{\text{dec}} 5$ units ✓ more change

$x_3 \xrightarrow{inc}$ 1 unit, $y \xrightarrow{dec}$ 4 unit

feature imp of $x_i \rightarrow |w_i|$

$y = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_{18} x_{18}$

y (resale price) $= w_0 + w_1$ (age) $+ w_2$ (km_driven)

$w_1 \rightarrow \boxed{10{,}000}$, $w_2 \rightarrow 100$

age $\rightarrow$ 0 to 50 years

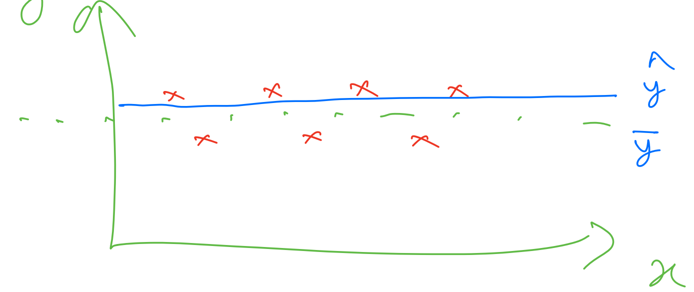
km_driven $\rightarrow$ 0 to 1,00,000 km

Scaling $\rightarrow$ zero mean, unit std

Standard scaler

min max scaler

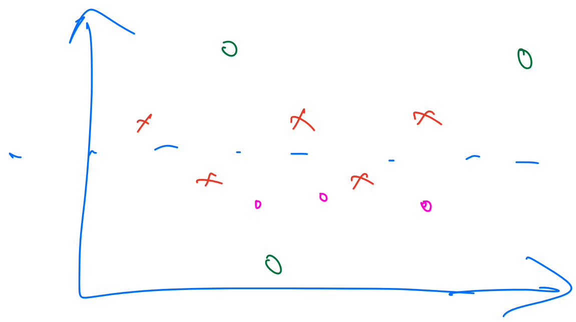
model interpretability



$\boxed{MSE \rightarrow 0.5}$

$\boxed{RMSE \rightarrow 0.43}$

$\boxed{0.5,\ 0.6,\ 0.7}$



$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

with Outliers → 0.5, 100

w/o outliers → 0.5