Linear Regression.

Agenda:

1. Intr. to ML. → example.
2. Formulation of LR. →
3. optimization concept
4. Solution of LR.
5. codenotebook of LR.

(*). House Rent Prediction:-

house → Model → Rent

Feature eng.

#rooms
landmarks
Amenities
Area
Fur/not
Age
Economic factors

Rent

↳ 2 features that represent home → [#rooms, # landmarks] Rent

I/P

$\begin{bmatrix} 2 \\ 5 \end{bmatrix}$

Rule

Rent = 1000 × #rooms + 500 × # landmarks → 1000×2 + 500×5 = 4500

Problem with model!

↳ No backing of "DATA"

"Intuit. backed up by data"

Training Data:

[5] [9]

I/P (house)	O/P (Rent)
H1 → [2, 5]	20k
H2 → [3, 6]	30k
H3 → [4, 8]	
H10 → [7, 6]	

$2 \times 1000 + 5 \times 500 = 4500$

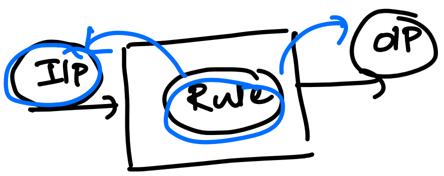
far apart.

$3 \times 1000 + 6 \times 500 = 6000$

w & m

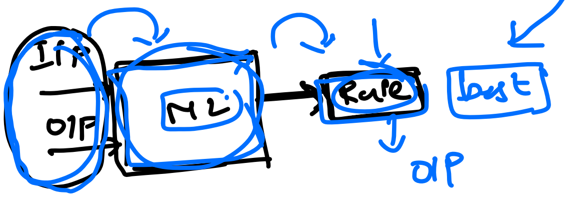
$w = m \cdot g$

Traditional p-way:



ML paradigm

new I/P



$\text{Rent} = \underbrace{1000}_{w_1} \times \#\text{rooms} + \underbrace{500}_{w_2} \times \#\text{bedrooms} + w_3$

Not "RESPECTS" data.

[*] our obj is to find best values of w_1, w_2, w_3 that Respects data

---

Formulate:

$$\mathcal{D} = \{ (x_1, y_1), (x_2, y_2), \dots (x_m, y_m) \}$$

x: I/P: vector of 'n' features y: Actual o/p

m: # datapts

n: # features eg. o/p.

$$\text{Rule } \underbrace{\hat{y}}_{\text{predicted O/P}} = w_1 \times f_1 + w_2 \times f_2 + \dots + w_n \times f_n + w_{n+1}$$

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}_{n \times 1}$$

$$\mathcal{D} = \{ ([2,5], 20k), ([3,6], 30k), \dots \}$$

f_1: #rooms

f_2: #bedrooms

m: # friends

n: 2

$\begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}_{n \times 1}$

$D = \{(x_1, y_1), \ldots (x_m, y_m)\}$

$\hat{y} = \boxed{w_1} f_1 + w_2 \cdot f_2 + \cdots + w_n f_n + \boxed{w_{n+1}}$

$D = \{(x_1, y_1), \ldots (x_m, y_m)\}$

1-dim vector

$\hat{y} = mx + c$

$w_1 f_1 + w_2$

$w_1 x + w_2$

$\hat{y} = w_1 \times \boxed{\#rooms} + w_2 \times$ # ladder $+ w_3 \ldots$

Find/learn the best/optimal value of $w_1, w_2, \ldots w_{n+1}$ that 'respect' the data.

"ensure the diff b/w pred. o/p & actual o/p is as small as possible"

Loss fn / error fn) = (Actual o/p − predic o/p)^2.

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

{ obj is to learn optimal value of $w_1, w_2, \ldots w_{n+1}$ that MINIMIZES $\sum_{i=1}^{m} (y_i - \hat{y}_i)^2$ } 'Respect'

Act. Pred.

2k 2k

20k

18k 22k

$|y_i - \hat{y}_i|$

$(y_i - \hat{y}_i)^2$ — math. conv. 30m

objective is to find value of $w_1 \ldots w_{n+1}$

that minimize $\left\{ \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \right\}$

$\hat{y}_i = w_1 \times f_1 + w_2 \times f_2 + \ldots w_n \times f_n + w_{n+1}$.

$\min g(w)$

minimize $g(w)$

7:14 2→ ___

---

minimize a function $g(w)$ w:

* under some assumption on g → CONVEX

I. Compute derivative of fn. $g'(w)$

II. Equate it to zero. $w \rightarrow g'(w) = 0$

$w = 1$

Ex1: $g_1(w) = (w-1)^2 + 2.$

w

I. $g_1'(w) = 2(w-1)$

II. $2(w-1) = 0 \Rightarrow w = 1$ ✓

$g_2(w) = (w-3)^2 + 5.$ $(w=3)$.

I. $g_2'(w) = 2(w-3)$

II. $2(w-3) = 0 \Rightarrow w = 3$

$g(x) = |x| \rightarrow$ not diff. at 0 ?

$g \;\; x^2 \rightarrow$ diff.

$\boxed{g(w_1, w_2, \ldots w_n)} \rightarrow$

$$g(w_1, w_2) = 2w_1^2 + 3w_1w_2 + 4w_2^2$$

Partial derivative :

$$\nabla g = \begin{bmatrix} \partial g / \partial w_1 \\ \partial g / \partial w_2 \end{bmatrix} = \begin{bmatrix} 4w_1 + 3w_2 \\ 3w_1 + 8w_2 \end{bmatrix}$$

Rule book for finding minimum of multivariate : $g(w_1, w_2, \ldots w_n)$.

I. compute gradient $\nabla g = \begin{bmatrix} \partial g / \partial w_1 \\ \partial g / \partial w_2 \\ \vdots \\ \partial g / \partial w_n \end{bmatrix}_{n \times 1}$

II. equate to 0. $\boxed{\nabla g = 0.}$ $w_1, w_2, \ldots w_n$

Applying this rulebook on $g(w_1, \ldots w_n) = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$ $\cancel{w_{n \times 1}}$

where $\hat{y}_i = w_1 \cdot f_1 + w_2 \cdot f_2 + \ldots w_n \cdot f_n$

3 Key matrices of linear Re:- (Simple) :-

m: #

n:

$X \rightarrow$ i/p Matrix

$Y \rightarrow$ o/p matrix

$W \rightarrow$ weight matrix

$$X = \begin{array}{c} 1 \\ 2 \\ m \end{array} \begin{bmatrix} f_1 & \cdots & f_n \\ f_1 & \cdots & f_n \\ f_1 & \cdots & f_n \end{bmatrix}_{m \times n} \quad W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{m \times 1}$$

$$\begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$D = \{ ([2,5], 20k), ([3,6], 30k), ([4,7], 40k) \}$

$$X = \begin{bmatrix} 2 & 5 \\ 3 & 6 \\ 4 & 7 \end{bmatrix}_{3 \times 2}$$

$$\vec{W} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}_{2 \times 1}$$

$$Y = \begin{bmatrix} 20k \\ 30k \\ 40k \end{bmatrix}$$

$$J(w_1 \ldots w_n) = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

$$\hat{y}_i = w_1 \cdot f_1 + w_2 \cdot f_2 + \ldots + w_n \cdot f_n \cdot$$

$X_{m \times n}$, $Y_{m \times 1}$, $W_{n \times 1}$.

$$X \cdot W = \begin{bmatrix} f_1 & \ldots & f_n \\ f_1 & \ldots & f_n \\ f_1 & \ldots & f_n \end{bmatrix}_{m \times n} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}_{n \times 1}$$

$$X \cdot W = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_m \end{bmatrix}_{m \times 1} = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}_{m \times 1}$$

$\hat{y}_2$, $\hat{y}_2$, $\hat{y}_m$

$$\boxed{Y - XW} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_m \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_m - \hat{y}_m \end{bmatrix}_{m \times 1}.$$

$$\sum_{i=1}^{m} (y_i - \hat{y}_i)$$

$$(y - Xw)^T (y - Xw)$$

$u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}_{m \times 1}$

$u^T u = [u_1 \; \cdots \; u_m] \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = u_1^2 + u_2^2 + \cdots + u_m^2$

$$g(w_1, \ldots w_n) = (y - Xw)^T (y - Xw)$$

$(y - Xw)^2$

I. compute gradient of g. ∇g

$\begin{bmatrix} \partial g / \partial w_1 \\ \vdots \\ \partial g / \partial w_n \end{bmatrix} \leftarrow \underset{n \times 1}{\nabla g} = -2X^T \underset{m \times 1}{(y - Xw)}$

eg things like derivative write like gradient as

$-2(y - Xw)X$

$X : m \times n$

$X^T : n \times m$

$u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

$u^T u = [1 \; 2 \; 3] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1^2 + 2^2 + 3^2$

$$g = (y - Xw)^T (y - Xw).$$

$$\nabla g = \begin{bmatrix} \partial g / \partial w_1 \\ \vdots \\ \partial g / \partial w_n \end{bmatrix}_{n \times 1} = -2 \underset{n \times m}{X^T} \underset{m \times 1}{(y - Xw)}$$

$(y - Xw)^2 \rightarrow -2(y - Xw)X$

$X : m \times n$

$m \times 1$

$$\rightarrow \boxed{\nabla g = -2X^T(y - Xw)}$$

II. $\nabla g = 0 \Rightarrow -\cancel{2}X^T(y - Xw) = 0$

$\Rightarrow X^T y - X^T X w = 0$

$$\Rightarrow (X^T X) w = X^T y.$$

$$\Rightarrow w = (X^T X)^{-1} X^T y.$$

$n \times 1$