Univariate linear regression

Size in feet² (x)	Price ($) in 1000's ($y$)
2104	400
1416	232
1534	315
852	178
...	...

← output

↑ one features

features (inputs) | outputs

Size in feet² x_1	Number of bedrooms x_2	Number of floors x_3	Age of home in years x_4	Price ($) in $1000's
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...	...	...	...	...

i = 2

col

$j = 1 \ldots 4$

$x_1^{(2)} = 1416$ (row: 2, column: 1)

x_j = j^{th} feature

n = number of features ($n = 4$ here)

$\vec{x}^{(i)}$ = features of i^{th} training example

If we talk about the whole row,

$\vec{x}^{(2)} = [1416 \quad 3 \quad 2 \quad 40]$ → row vector

$x_1^{(2)}$ $x_2^{(2)}$ $x_3^{(2)}$ $x_4^{(2)}$

Previously, the model was :-

$$f_{w,b}(x) = wx + b$$

↑ only one variable

Currently,

$$f_{\vec{w},b}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b$$

w and x are vectors

example

$$f_{w,b}(x) = 0.1x_1 + 4x_2 + 10x_3 - 2x_4 + 80$$

x_1 ↑ size

x_2 ↑ no. of bedrooms

x_3 ↓ no. of floors

x_4 ↑ years

80 ↓ base price

1. size – For every increase in some size of the house, the price will also increase.

2. no. of bedrooms – The more the no. of bedrooms the more the price of the house.

3. no. of floors – The more the no. of floors the more the price.

4. no. of years - The older the house, the lesser the price.

$$f_{\vec{w},b}(\vec{x}) = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots + w_n x_n$$

$\vec{w} = [w_1 \quad w_2 \quad w_3 \quad \ldots \quad w_n]$

b is a number (in above context, \$80K is the base, ⇒ all houses are above 80K)

parameters

$\vec{x} = [x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n]$

dot product → $\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i$

$$f_{w,b}(x) = \vec{w} \cdot \vec{x} + b = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots + w_n x_n + b$$

The above type of linear regression isn't called multivariate regression because that refers to a different type of regression. We'll refer to this type of algorithm to simply as multiple linear regression.