3. Linear Regression

Consider the given data

Response Variable (y)

Regressor / Explanatory variable (x)

Year	Sales (₹)	Advertising (₹)
1	651	23
2	762	26
3	856	30
4	1063	34
5	1190	43
6	1298	48
7	1421	52
8	1440	57
9	1518	58

Q: How to predict the Sales, as a function of Advertising?

Regression.

Regression is an ML technique, which precisely addresses this problem.

Regression : Algorithm to predict a Response variable, based on a set of Regressors or Explanatory variable.

Multiple Linear regression

	TV (x_1)	Radio (x_2)	Newspaper (x_3)	Sales (y)
0	230.1	37.8	69.2	922.1
1	44.5	39.3	45.1	510.4
2	17.2	45.9	69.3	129.3
3	151.5	41.3	58.5	618.5
4	180.8	10.8	58.4	712.9

Regressors / Explanatory variables ($\vec{x}$)

Response variable (y).

In general, the regressor $\bar{x}$ can be an n-dimensional vector.

* $x_1 \rightarrow$ cost of TV advertising
* $x_2 \rightarrow$ cost of Radio advertising
* $x_3 \rightarrow$ cost of Newspaper advertising and so on...

Other examples

(i) $y(k) \rightarrow$ Price of particular stock at time k

$$\left.\begin{matrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_N(k) \end{matrix}\right\} \rightarrow \text{Prices of related stocks at time } k$$

(ii) $y(k) \rightarrow$ Sales of SUV at time k

$$\left.\begin{matrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_N(k) \end{matrix}\right\} \rightarrow \text{Sales of bikes, Sales of cars, Average income, ... at time } k.$$

Linear Regression

The outputs $y(k)$ can be modeled using a linear combination of Regressors / Explanatory variables $x_i(k)$.

$$y(k) = h_0 + h_1 x_1(k) + \dots + h_n x_n(k) + \epsilon(k)$$

$$= \begin{bmatrix} 1 & x_1(k) & x_2(k) & \dots & x_n(k) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_n \end{bmatrix} + \epsilon(k)$$

$$= \bar{x}^T(k)\,\bar{h} + \epsilon(k)$$

This is termed Linear Regression.

$h_0, h_1, \dots h_n$ are termed Regression coefficients

The Regression coefficients can be computed as follows.

- Consider the availability of training pairs $(y(k), \bar{x}(k))$ for $k = 1, 2, \ldots, M$.
- The <u>training set</u> can be expressed as

$$\underbrace{\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(M) \end{bmatrix}}_{\bar{y}} = \underbrace{\begin{bmatrix} \bar{x}^T(1) \\ \bar{x}^T(2) \\ \vdots \\ \bar{x}^T(M) \end{bmatrix}}_{X} \bar{h} + \underbrace{\begin{bmatrix} \epsilon(1) \\ \epsilon(2) \\ \vdots \\ \epsilon(M) \end{bmatrix}}_{\bar{\epsilon}}$$

- $\bar{y} = X\bar{h} + \bar{\epsilon}$

- X is a TALL matrix.

The matrix $(X^T X)^{-1} X^T$ is termed as the <u>pseudo-inverse</u> of X, since

$$\underbrace{(X^T X)^{-1} X^T}_{\text{pseudo-inverse}} \times X = I$$

- Recall, the learning model is $\bar{y} = X\bar{h} + \bar{\epsilon}$

To determine regression coefficients $\bar{h}$, solve the problem

$$\min \| \underbrace{\bar{y} - X\bar{h}}_{\bar{\epsilon}} \|^2$$

This is termed as the <u>Least Squares</u> (LS) problem.

- The <u>Least squares</u> (LS) solution is

$$\min \| \bar{y} - X\bar{h} \|^2 = (\bar{y} - X\bar{h})^T (\bar{y} - X\bar{h})$$

$$= (\bar{y}^T - \bar{h}^T X^T)(\bar{y} - X\bar{h})$$

$$= \bar{y}^T\bar{y} - \bar{h}^T X^T \bar{y} - \bar{y}^T X \bar{h} + \bar{h}^T X^T X \bar{h}$$

⊙ To minimize, evaluate the gradient w.r.t. $\bar{h}$ and set it equal to zero.

The gradient w.r.t $\bar{h}$ is defined as

$$\nabla_f(\bar{h}) = \begin{bmatrix} \frac{\partial f}{\partial h_1}(\bar{h}) \\ \frac{\partial f}{\partial h_2}(\bar{h}) \\ \vdots \\ \frac{\partial f}{\partial h_t}(\bar{h}) \end{bmatrix}$$

⊙ Use the principles

* $\nabla \bar{c}^T \bar{h} = \bar{c}$

* If $P^T = P$, $\nabla \bar{h}^T P \bar{h} = 2 P \bar{h}$

⊙ Solve using Lagrangian

$$f(\bar{h}) = \bar{y}^T \bar{y} - \bar{h}^T X^T \bar{y} - \bar{y}^T X \bar{h} + \bar{h}^T X^T X \bar{h}$$

$$\nabla_f(\bar{h}) = -2 X^T \bar{y} + 2 X^T X \bar{h} = 0$$

$$\Rightarrow \boxed{\bar{h} = (X^T X)^{-1} X^T \bar{y}}$$

This is the Least Squares (LS) solution.

Problem 1 : Calculate $(X^TX)^{-1}$, given $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{\det(X^TX)} \operatorname{adj}(X^TX)$$

$$= \frac{1}{120-100} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

Problem 2 : Find the pseudo-inverse of $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$

$$\text{Pseudo-inverse of } X = (X^TX)^{-1} X^T$$

$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 20 & 10 & 0 & -10 \\ -6 & -2 & 2 & 6 \end{bmatrix}$$

Problem 3 : Solve the LS problem below and calculate $\bar{h}$

$$\min \left\| \underbrace{\begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}}_{\bar{y}} - \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} h_0 \\ h_1 \end{bmatrix}}_{\bar{h}} \right\|^2$$

2 Regression coefficients

M=4
Training samples

Rows > # columns
(ie) 4 > 2
⇒ Tall Matrix

We have, $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, $\bar{y} = \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$, $\bar{h} = \begin{bmatrix} h_0 \\ h_1 \end{bmatrix}$

$$\bar{h} = (X^T X)^{-1} X^T \cdot \bar{y}$$

$$= \frac{1}{20} \begin{bmatrix} 20 & 10 & 0 & -10 \\ -6 & -2 & 2 & 6 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -3 \\ 1 \end{bmatrix}$$

$$\boxed{\bar{h} = \frac{1}{20} \begin{bmatrix} -10 \\ 2 \end{bmatrix}}$$

This is the Regression coefficient vector.

(ii) For a given set of training data, and given a set of inputs and outputs; this is how we learn the Regression coefficients.