Arif Romadhan
arifromadhan19@gmail.com

Linear Regression

Basic of Simple and Multiple Linear Regression

1. Introduction

Linear regression is the supervised learning approach. In particular, linear regression is a useful method for predicting continuous values (target) and attempts to model the linear relationship between target and one or more predictor.

2. Simple Linear Regression (predictor = 1)

Simple linear regression lives up to its name: simply, finding a relationship between one predictor (X) and target (Y). Mathematically, we can write this linear relationship as

$\hat{y} = \widehat{f}(x)$ (1.0)
$\hat{y} = {}_{0} + {}_{1}$ (1.1)

In Equation 1, $\hat{y}$ is target, is Predictor, ${}_{0}$ and ${}_{1}$ (coefficients) are two unknown constants that represent the intercept and slope terms in the linear model. Simple linear regression attempts to produce estimates ${}_{0}$ and ${}_{1}$ for the model coefficients. ${}_{0}$ and ${}_{1}$ obtained, we can predict the target.

2.1 Residual

Linear regression actual Response

$\hat{y} = \widehat{f}(x) + \varepsilon$ (2.0)
$\hat{y} = {}_{0} + {}_{1} + \varepsilon$ (2.1)

ε is residual error. Mathematically, we can write residual error in linear regression as following

$$\varepsilon_i = \ _i - \hat{y}_i \quad (3.0)$$

this is the difference between the i-th actual value and the i-th predicted value. We define the RSS (residual sum of squares) as

$$RSS = \sum_{i=1}^{n} \varepsilon_i^2 \quad (4.0)$$

$$RSS = \varepsilon_1^2 + \varepsilon_2^2 + .. + \varepsilon_n^2 \quad (4.1)$$

or equivalently as

$$RSS = (\ _1 - \ _0 - \ _1 x_1)^2 + (\ _2 - \ _0 - \ _1 x_2)^2 + .. + (\ _n - \ _0 - \ _1 x_n)^2 \quad (4.2)$$

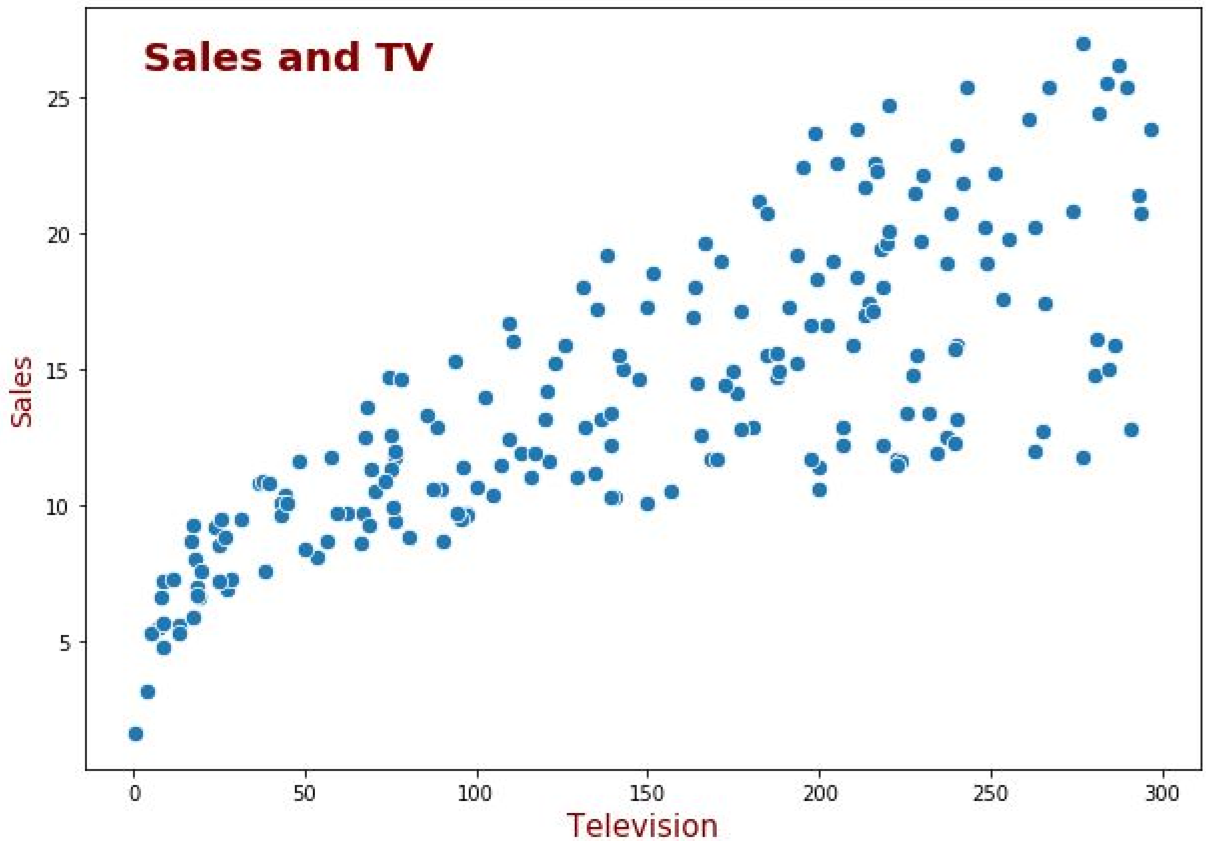
2.2 Estimating the Coefficients

$_0$ and $_1$ are unknown. So before we can use equation 1.1 to make predictions, we must use data to estimate the coefficients. As mentioned in equation 4.2, the least squares approach chooses $_0$ and $_1$ to minimize the RSS. Using some calculus $_0$ and $_1$ can we write as

$$_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (5.0)$$

$$_0 = \bar{y} + \ _1 \bar{x} \quad (6.0)$$

Where $\bar{y}$ is mean of y (target) and $\bar{x}$ is mean of (predictor). In practice, we use Advertising case within TV, Radio, Newspaper as predictor and sales as target. In order to implementation simple linear regression, The only TV predictor will be used in practice. Graph of sales and tv as following :



2.3 Implementation

As mentioned above, we only use TV as (predictor), Sales as y (target) and number of data is 200 row (github repository)

	x	y	pow((x-x_mean),2)	(x-x_mean)	(y-y_mean)	(x-x_mean)(y-y_mean)
0	230.1	22.1	6898.55	83.06	8.08	671.1248
1	44.5	10.4	10514.96	-102.54	-3.62	371.1948
2	17.2	9.3	16859.07	-129.84	-4.72	612.8448
3	151.5	18.5	19.87	4.46	4.48	19.9808
4	180.8	12.9	1139.57	33.76	-1.12	-37.8112
5	8.7	7.2	19138.65	-138.34	-6.82	943.4788
6	57.5	11.8	8017.86	-89.54	-2.22	198.7788
7	120.2	13.2	720.52	-26.84	-0.82	22.0088
8	8.6	4.8	19166.33	-138.44	-9.22	1276.4168
9	199.8	10.6	2783.35	52.76	-3.42	-180.4392

First, we solve for the coefficient $_1$ (slope) with $\overline{y}$ = 14.0225 and $\overline{x}$ = 147.0425

$$_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

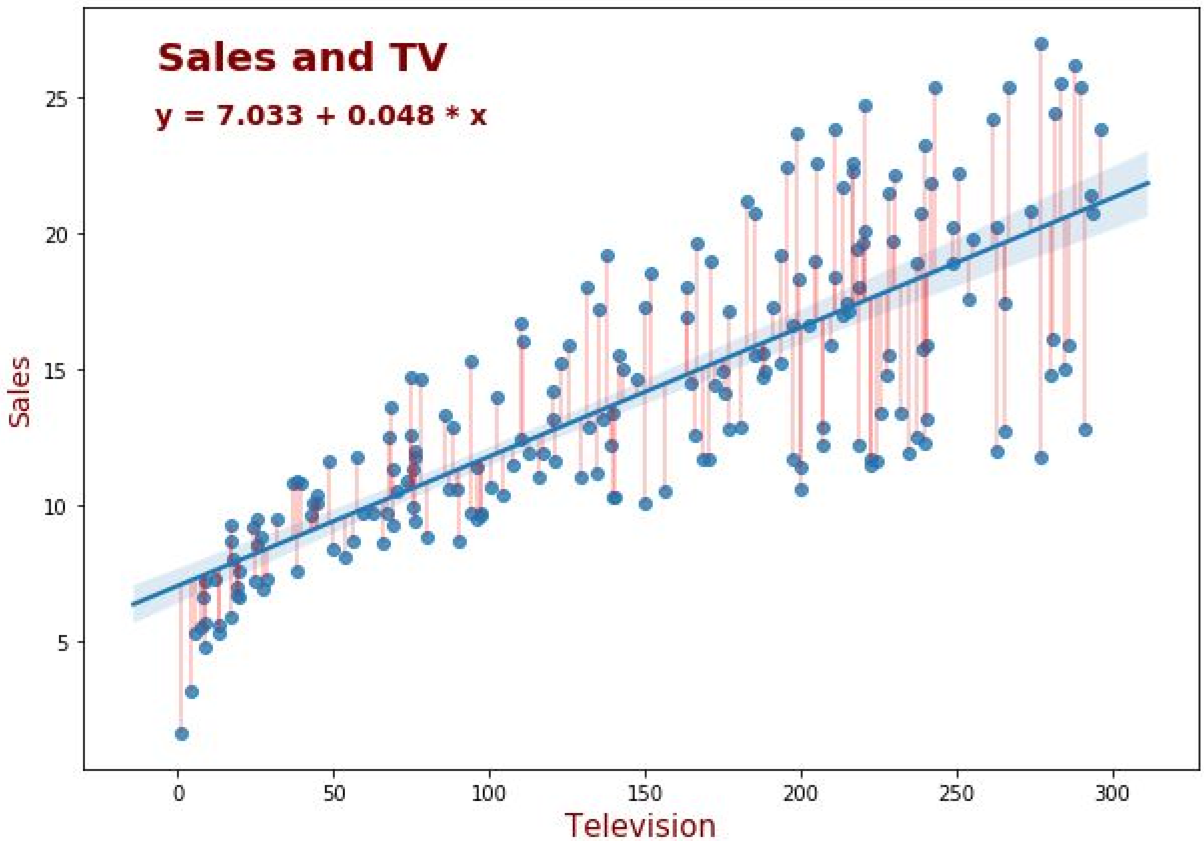
$_1$ = 69727.65 / 1466818.94
$_1$ = 0.047536644161412324
$_1$ = 0.048

Once we know the value of the coefficient $_1$, we can solve for the coefficient $_0$ (intercept) :

$_0 = \bar{y} + {}_1\bar{x}$
$_0$ = 14.022 - $_1$ * 147.04
$_0$ = 7.033

Therefore, the regression equation is
ŷ = 7.033 + (0.048 *)

Regression graph of sales and TV as following :



2.4 How to Use the Regression Equation

since we have regression equation 1.1, we can predict sales for new data (tv). In our example, new data of tv is 100, the predicting sales with 100 tv as following

$\hat{y}$ $= {}_0 + {}_1$
$\hat{y}$ = 7.033 + (0.048 * new_data)
$\hat{y}$ = 7.033 + (0.048 * 100)
$\hat{y}$ = 11.833
Sales = 11.833

3. Multiple Linear Regression (predictor > 1)

Dealing with such a large number of values and operations computers tend to perform efficiently matrix, To express the regression equation in matrix form, we need to define three matrices: , , and x.

$$
y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, \quad
x = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & . & . & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & x_{n,1} & x_{n,2} & . & . & x_{n,p} \end{bmatrix}, \quad
\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_n \end{bmatrix}, \quad
\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}
$$

General parametric equation :

$\hat{y} = \ x + \varepsilon$ (7.0)

$\hat{y} = {}_0 + {}_1x_1 + {}_2x_2 + ... + {}_nx_n + \varepsilon$ (7.1)

Representation equation 7.0 in matrix form

$$
\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} =
\begin{bmatrix}
\beta_0 & + & \beta_1 x_{1,1} & + & \beta_2 x_{1,2} & + & . & . & + & \beta_p x_{1,p} & + & \varepsilon_1 \\
\beta_0 & + & \beta_1 x_{2,1} & + & \beta_2 x_{2,2} & + & . & . & + & \beta_p x_{2,p} & + & \varepsilon_2 \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
\beta_0 & + & \beta_1 x_{n,1} & + & \beta_2 x_{n,2} & + & . & . & + & \beta_p x_{n,p} & + & \varepsilon_n
\end{bmatrix}
$$

3.1 Residual

Similar to simple linear regression, multiple linear also calculates the residual error represented in the form of a matrix

$$
\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix} =
\begin{bmatrix} y_1 - \hat{y_1} \\ y_2 - \hat{y_2} \\ . \\ . \\ y_n - \hat{y_n} \end{bmatrix} =
\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} -
\begin{bmatrix} \hat{y_1} \\ \hat{y_2} \\ . \\ . \\ \hat{y_n} \end{bmatrix} = y - \hat{y}
$$

RSS in equation 4.0 can be re-written as

$RSS = \sum_{i=1}^{n} \varepsilon_i^2 \Rightarrow \varepsilon^T \varepsilon$ (8.0)

$RSS = (\; - \hat{y})^T(\; - \hat{y})$ (8.1)
$RSS = (\; - x \;)^T(\; - x \;)$ (8.2)
$RSS = (\; - x^{T\,T})(\; - x \;)$ (8.3)
$RSS = \; ^T \; - \; ^T x \; - x^T \; + \; ^T x^T x$ (8.4)

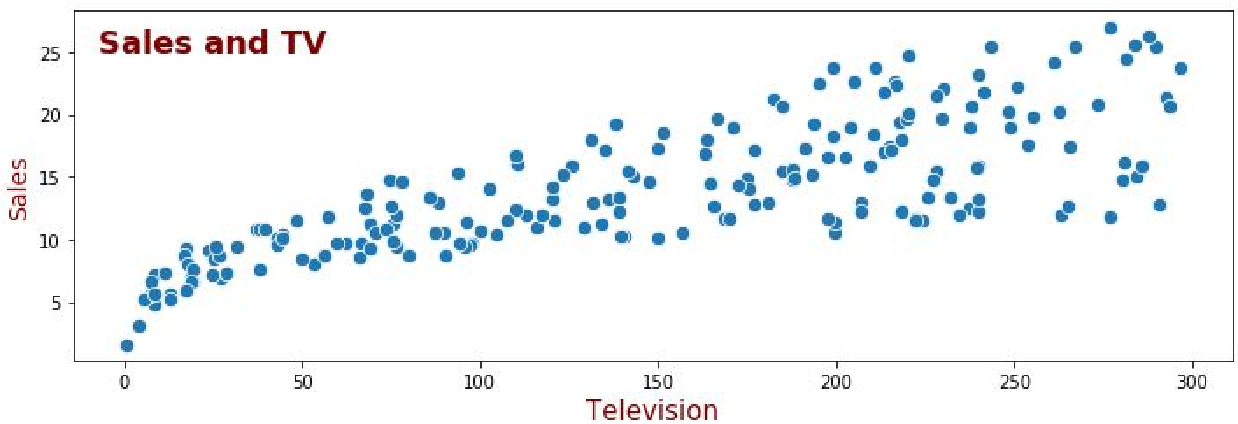
3.2 Estimating the Coefficients

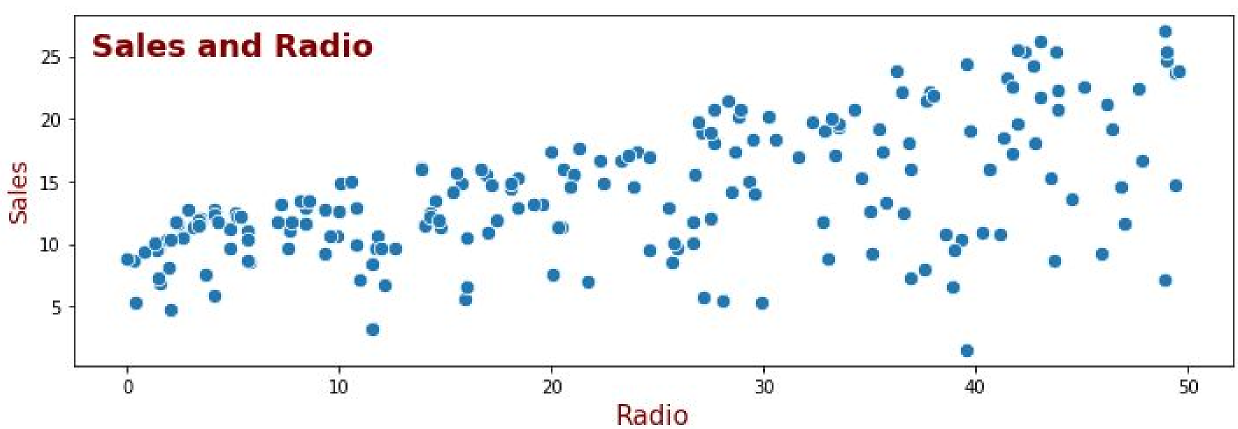
Computing coefficients in multiple regression are the same thing in simple linear regression, where minimizing the RSS (residual sum of squares). Using some calculus Computing coefficients can be written as

$= (x^T x \;)^{-1} x^T$ (9.0)

3.3 Implementation

same as simple linear regression, we also use Advertising case within TV and Radio as predictors and sales as target. Graph of tv, radio, and sales as following :





What the next? We define x , x^T , compute $x^T x$, find the inverse $x^T x$, and define from this table where TV as x_1 , Radio as x_2 and Sales as

	TV	Radio	Sales
0	230.1	37.8	22.1
1	44.5	39.3	10.4
2	17.2	45.9	9.3
3	151.5	41.3	18.5
4	180.8	10.8	12.9

	x1	x2	y
0	230.1	37.8	22.1
1	44.5	39.3	10.4
2	17.2	45.9	9.3
3	151.5	41.3	18.5
4	180.8	10.8	12.9

$$x = \begin{bmatrix} 1 & 230.1 & 37.8 \\ 1 & 44.5 & 39.3 \\ 1 & 17.2 & 45.9 \\ 1 & 151.5 & 41.3 \\ 1 & 180.8 & 10.8 \end{bmatrix}, \quad x^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 230.1 & 44.5 & 17.2 & 151.5 & 180.8 \\ 37.8 & 39.3 & 45.9 & 41.3 & 10.8 \end{bmatrix}, \quad y = \begin{bmatrix} 22.1 \\ 10.4 \\ 9.3 \\ 18.5 \\ 12.9 \end{bmatrix}$$

$$x^T x = \begin{bmatrix} 5 & 624.1 & 175.1 \\ 624.1 & 110862 & 19445.7 \\ 17.5 & 19445.7 & 690.2 \end{bmatrix}, \quad (x^T x)^{-1} = \begin{bmatrix} 3.9 & -0.0092 & -0.0743 \\ -0.0092 & 393349 & 12305 \\ -0.0743 & 0.0001 & 0.0016 \end{bmatrix}$$

After we have x , x^T , $x^T x$, inverse $x^T x$, and . we can use equation 9.0

$= (x^T x)^{-1} x^T$

$$\beta_0 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -1.898 \\ 0.068 \\ 0.228 \end{bmatrix}$$

3.4 How to Use the Regression Equation

since we have regression equation 7.0, we can predict sales for new data (TV and Radio). In our example, new data of tv is 100 and radio is 200, the predicting sales with 100 tvs as following

ŷ = $_0 + {}_1x_1 + {}_2x_2$

ŷ = -1.898 + (0.068* new_data TV) + (0.228* new_data Radio)

ŷ = -1.898 + (0.068 * 100) + (0.228 * 200)

ŷ = 50.502

Sales = 50.502

Reference

1. Introduction to Statistical Learning
2. https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86
3. https://stattrek.com/multiple-regression/regression-coefficients.aspx
4. https://www.youtube.com/watch?v=K_EH2abOp00

About Me

I'm a Data Scientist, Focus on Machine Learning and Deep Learning. You can reach me from :

Email : **arifromadhan19@gmail.com**
LinkedIn : **https://www.linkedin.com/in/arif-romadhan-292116138/**
Github : **https://github.com/arifromadhan19/linear_regression**