$Linear\ regression\ Session\ -\ 2$

$Linear\ regression\ equation:\ \ y\ =\ b_o\ +\ b_1\ *\ x_1\ +\ b_2\ *\ x_2\ +\ \dots\ +\ b_n\ *\ x_n$

- $Generally\ \ we\ have\ input\ column\ and\ we\ also\ have\ target\ data$
- $input\ columns\ we\ denoted\ with\ x_1,\ x_2,\ \dots x_n$
- $output\ column\ or\ target\ column\ denoted\ with\ y$
- $this\ target\ column\ y\ ,\ we\ called\ actual\ output:\ y_{actual}$
- $we\ have\ one\ more\ name:\ Ground\ truth\ data$
- $model\ will\ train\ by\ passing\ input\ data\ and\ actual\ output\ data$
- $Once\ model\ is\ developed\ we\ will\ pass\ again\ input\ data\ only$
- $so\ model\ also\ will\ give\ some\ outputs,\ this\ output\ is\ called\ as\ predictions:\ y_{predictions}$
- $in\ order\ to\ evaluate\ the\ model\ performance\ we\ need\ to\ comapre\ y_{actual} with\ y_{prediction}$

$100, 50, 100, 50:\ assumption$

$$Sales\ =\ b_o\ +\ b_1\ *\ SM\ +\ b_2\ *\ TV\ +\ b_3\ *\ NP$$

$$Sales\ =\ 1\ +\ 5\ *\ SM\ +\ 2\ *\ TV\ +\ 2\ *\ NP$$

SM	TV	NP	$Sales\ =\ y\_actu$	$sales\ =\ y\_predic$	Error: $y_{actual}\ -\ y\_pre$
300	400	500	1000	80100	Error1
200	150	500	2000	50100	Error2
100	200	300	500	40100	Error3
400	500	200	3000	80100	Error4
300	150	150	1000	37600	Error5

$$100\ +\ 50\ *\ 300\ +\ 100\ *\ 400\ +\ 50\ *\ 500$$

$$100\ +\ 50\ *\ 200\ +\ 100\ *\ 150\ +\ 50\ *\ 500$$

$$100\ +\ 50\ *\ 100\ +\ 100\ *\ 200\ +\ 50\ *\ 300$$

$$100\ +\ 50\ *\ 400\ +\ 100\ *\ 500\ +\ 50\ *\ 200$$

$Error:$

$Error = \left(y_{actual} - y_{prediction}\right) \;\; or \;\; (y_a - y_p)$

- $for\ every\ observations\ has\ actual\ output,$
- $for\ every\ observation\ model\ will\ give\ prediction$
- $will\ compare\ y_{actaul}\ and\ y_{prediction}\ observation\ by\ observation$

$$Error_1 = (y_{a_1} - y_{p_1})$$

$$Error_2 = (y_{a_2} - y_{p_2})$$

$$Error_3 = (y_{a_3} - y_{p_3})$$

$$Error_4 = (y_{a_4} - y_{p_4})$$

$$Error_5 = (y_{a_5} - y_{p_5})$$

$Total\ error$

$$TE = e_1 + e_2 + e_3 + e_4 + e_5$$

$$TE = \left(y_{a_1} - y_{p_1}\right) + \left(y_{a_2} - y_{p_2}\right) + \left(y_{a_3} - y_{p_3}\right) + \left(y_{a_4} - y_{p_4}\right) + (y_{a_5} - y_{p_5})$$

$$TE = \sum_{i=1}^{5}\left(y_{a_i} - y_{p_i}\right)$$

$$TE = e_1 + e_2 + e_3 \ldots e_n$$

$$TE = \left(y_{a_1} - y_{p_1}\right) + \left(y_{a_2} - y_{p_2}\right) + \left(y_{a_3} - y_{p_3}\right) + \;....+ \left(y_{a_n} - y_{p_n}\right)$$

$$TE = \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)$$

$Problem\ of\ Total\ Error$:

- $Total\ error\ is\ a\ summation\ of\ all\ individual\ errors$
- $one\ error\ might\ be\ positive\ and\ another\ error\ might\ be\ negative$
- $when\ we\ do\ sum\ of\ some\ postive\ values\ and\ some\ negative\ values$
- $there\ might\ be\ a\ chance\ the\ total\ error\ becomes\ zero$
- $we\ are\ seeing\ individual\ errors\ ,\ but\ total\ error\ zero\ where\ math\ fails$
- $this\ is\ same\ anlaogy\ of\ statistics\ mean\ diviation\ part$

$In\ order\ to\ avoid\ we\ need\ to\ do\ square\ of\ the\ errors$:

$SUM\ OF\ SQUARE\ ERRORS\ (SSE)$

$ERRORS$:

$$Error_1 = (y_{a_1} - y_{p_1})$$

$$Error_2 = (y_{a_2} - y_{p_2})$$

$$Error_3 = (y_{a_3} - y_{p_3})$$

$$Error_4 = (y_{a_4} - y_{p_4})$$

$$Error_5 = (y_{a_5} - y_{p_5})$$

$SQUARE\ ERRORS$

$$\left(Error_1\right)^2 = \left(y_{a_1} - y_{p_1}\right)^2$$

$$\left(Error_2\right)^2 = \left(y_{a_2} - y_{p_2}\right)^2$$

$$\left(Error_3\right)^2 = \left(y_{a_3} - y_{p_3}\right)^2$$

$$(Error_4 = \left(y_{a_4} - y_{p_4}\right)^2$$

$$\left(Error_5\right)^2 = \left(y_{a_5} - y_{p_5}\right)^2$$

$$SSE = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$

$$SSE = \left(y_{a_1} - y_{p_1}\right)^2 + \left(y_{a_2} - y_{p_2}\right)^2 + \left(y_{a_3} - y_{p_3}\right)^2 + \left(y_{a_4} - y_{p_4}\right)^2 + \left(y_{a_5} - y_{p_5}\right)^2$$

$$SSE = \sum_{i=1}^{5} \left(y_{a_i} - y_{p_i}\right)^2$$

$$SSE = \sum_{i=1}^{n} \left(y_{a_i} - y_{p_i}\right)^2$$

$MEAN\ SQUARE\ ERRORS\ (MSE)$

$$MSE = \frac{1}{n} * SSE$$

$$MSE = \frac{1}{n} * \sum_{i=1}^{n} \left(y_{a_i} - y_{p_i}\right)^2$$

$ROOT\ MEAN\ SQUARE\ ERRORS\ (RMSE)$

$$RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} \left(y_{a_i} - y_{p_i}\right)^2}$$

$$Error = (y_a - y_p)$$

$$error^2 = \left(y_a - y_p\right)^2$$

$$total\ error = \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)$$

$$SSE = \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)^2$$

$$MSE = \frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)^2$$

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)^2}$$

- *Error also called as Residual*
    - *SSE also called as Residual sum of Squares(RSS)*
- *Single error also called as Loss function*
- *Sum of square errors also called as : Residual sum of squares (RSS)*
- *Instead of* $e_1^2,\ e_2^2$ *you might have see* r_1^2, r_2^2
- *MSE also called as Cost function*
- *Cost function means sum of all errors (Losses)*
- *Cost function denoted with :* J

$$cost\ function = J = \frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)^2$$

Goal: we need to Find suitable coefficinets , in order to minmize the cost function

Suppose $f(x) = y$ *means, the equation y has only x variables*

$y = x + 2$

$f(x) = x + 2$

$f(x) = y$

similarly

$y = 2x + 3z$

$f(x, z) = 2x + 3z$

$f(x, z) = y$

$f(x, z)$ *means equation has combination of x and y*

$$cost\ function\ = J\ = \frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - y_{p_i}\right)^2$$

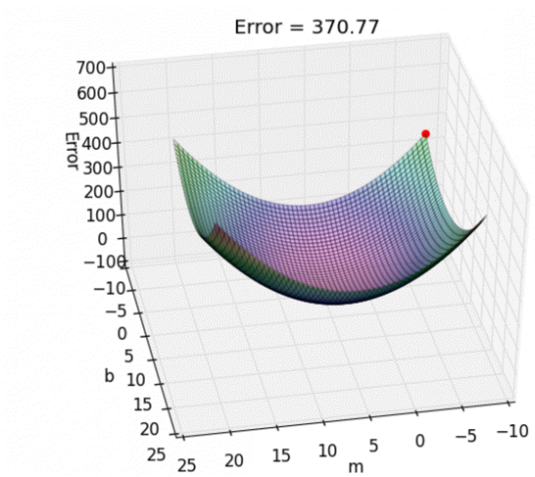
$$y_{p_i} = b_o + b_1 * x_1$$

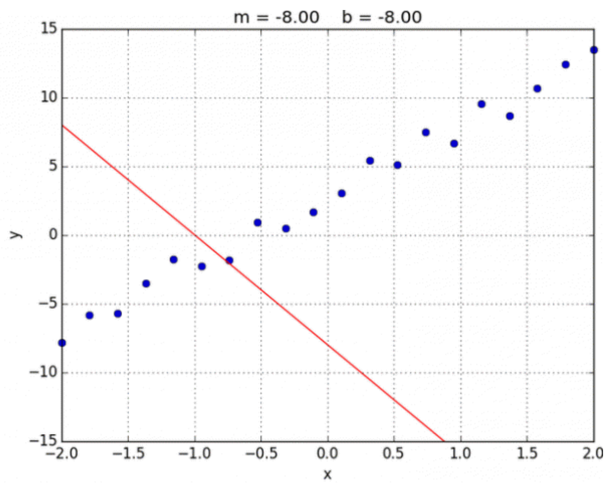
$$cost\ function\ = J(b_o, b_1)\ = \frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - (b_o + b_1 * x_1)\right)^2$$

$$cost\ function\ = J(b_o, b_1)\ = \frac{1}{n} * \sum_{i=1}^{n}\left(y_{a_i} - b_o - b_1 * x_1\right)^2$$

$$y_{a_i} = y^{\wedge}\ (hat\ means\ actual)$$

Goal: Find the suitable coefficients in order to Minimize the Cost function





Goal: Find the suitable coefficients in order to Minimize the Cost function

$$Minimum\ point\ means\ slope\ = 0$$

$$slope\ means\ = differentiation\ = \frac{dy}{dx}$$

$$so\ we\ need\ to\ differentiate\ the\ cost\ function\ (j) and\ makes\ equal\ to\ zero$$

$$but\ cost\ function\ has\ two\ parameters\ \left(b_o, b_1\right)$$

$$so\ we\ need\ to\ partial\ differentiation$$

$$Case\ -\ 1: \frac{\partial j}{\partial b_o} = 0\quad then\ we\ wil\ get\ \ b_o$$

$$Case\ -\ 2: \frac{\partial j}{\partial b_1} = 0\quad then\ we\ wil\ get\ \ b_1$$

$$The\ entire\ procedure\ known\ as\ OLS(ordinaty\ least\ square\ )Method$$