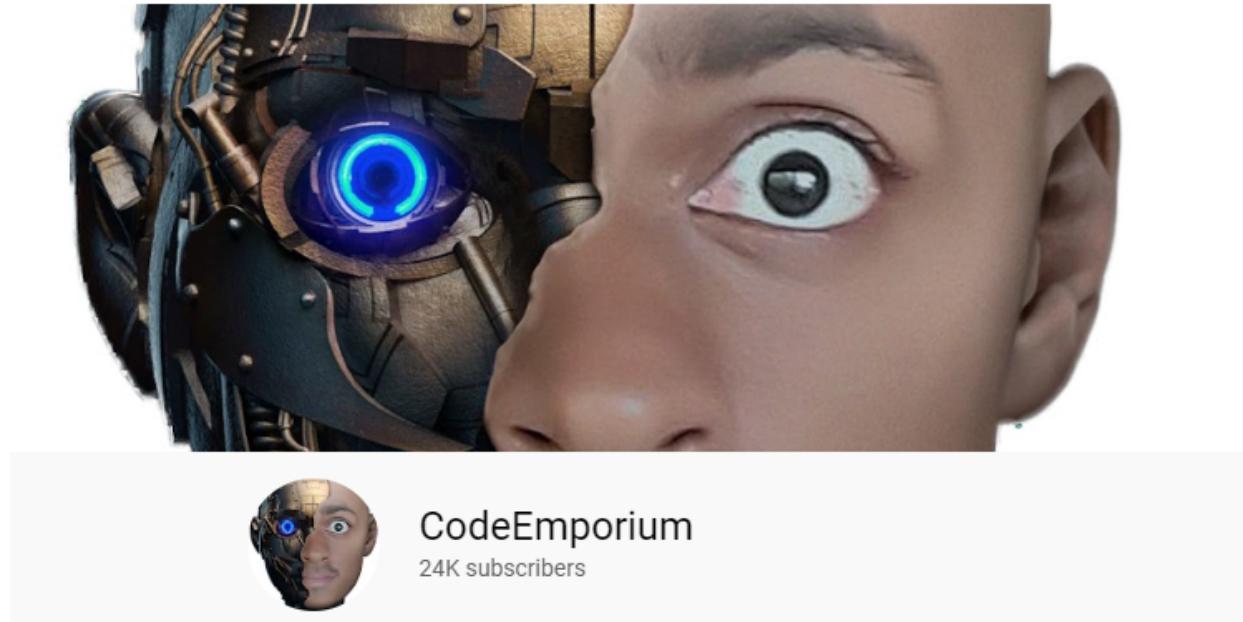


-->

thanks to code emporium for providing
wonderfull
tutorial and formulas and material in this video
some materials have been borrowed from code
emporium
so courtesy code emporium youtube channel

Figure 1:C:\Users\saktheeswaran\Desktop\jil.png



--> $y = f(x) + \epsilon;$

$$(\%o19) \quad y = \epsilon + f(x)$$

-->

$$--> \quad f(x) = \beta_0 + \beta_1 \cdot x;$$

$$(\%o20) \quad f(x) = \beta_1 x + \beta_0$$

-->

Figure 2:C:\Users\saktheeswaran\Desktop\linear regression in detail for deep learning\3est.png

$$\hat{y} = \widehat{f(X)}$$

-->

Figure 3:C:\Users\saktheeswaran\Desktop\linear regression in detail for deep learning\4est.png

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 X$$

Figure 4:C:\Users\saktheeswaran\Desktop\linear regression in detail for deep learning\7.png

$$\{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$$

Residual Error of sample i :

$$e_i = y_i - \hat{y}_i$$

Sum of Squared Residuals (RSS):

$$RSS = \sum_{i=1}^n e_i^2$$

Figure 5:C:\Users\saktheeswaran\Desktop\linear regression in detail for deep learning\8.png

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Figure 6:C:\Users\saktheeswaran\Desktop\linear regression in detail for deep learning\9.png

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_i))^2$$

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2$$

Figure 7: C:\Users\saktheeswaran\Desktop\bcap.png

$$\widehat{\beta}_1$$

--> ;

instead of beta cap using just beta a

hence the residual error from β cap changed to α which canm be written as

Figure 8:C:\Users\saktheeswaran\Desktop\linear regression in deteil for deep learning\8.png

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

--> argmin[$\beta[0], \beta[1]$]; sum(($y[i] - \alpha[i]$)², i, 1, n);

$$(\%o21) \quad \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \alpha_i)^2$$

now substitute α with $\alpha[0]$ and $\alpha[1]$

--> argmin[$\beta[0], \beta[1]$]; sum(($y[i] - (\alpha[0] + \alpha[1] \cdot x[i])$)², i, 1, n);

$$(\%o26) \quad \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)^2$$

--> diff(sum(($y[i] - \alpha[1] \cdot x[i] - \alpha[0]$)², i, 1, n), $\alpha[0]$, 1);

$$(\%o31) \quad -2 \sum_{i=1}^n y_i - \alpha_1 x_i - \alpha_0$$

--> -2 * sum($y[i] - \alpha[1] \cdot x[i] - \alpha[0]$, i, 1, n) = 0;

$$(\%o32) \quad -2 \sum_{i=1}^n y_i - \alpha_1 x_i - \alpha_0 = 0$$

-->

--> sum($y[i] - \alpha[1] \cdot x[i] - \alpha[0]$, i, 1, n);

$$(\%o33) \quad \sum_{i=1}^n y_i - \alpha_1 x_i - \alpha_0$$

--> $\text{sum}(y[i], i, 1, n) - \text{sum}(\alpha[1] \cdot x[i], i, 1, n) - \text{sum}(\alpha[0], i, 1, n);$

$$(\%o36) \quad -\alpha_0 n + \left(\sum_{i=1}^n y_i \right) - \alpha_1 \sum_{i=1}^n x_i$$

--> $\text{solve}(-\alpha[0] \cdot n + (\text{sum}(y[i], i, 1, n)) - \alpha[1] \cdot \text{sum}(x[i], i, 1, n), [\alpha[0]]);$

$$(\%o38) \quad [\alpha_0 = \frac{(\sum_{i=1}^n y_i) - \alpha_1 \sum_{i=1}^n x_i}{n}]$$

-->

--> $\alpha[0] = \text{sum}(y[i] - \alpha[1] \cdot x[i], i, 1, n) / n;$

$$(\%o39) \quad \alpha_0 = \frac{\sum_{i=1}^n y_i - \alpha_1 x_i}{n}$$

-->

Figure 9:

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

--> $\text{argmin}[\beta[0], \beta[1]] \cdot \text{sum}((y[i] - \alpha[i])^2, i, 1, n);$

$$(\%o40) \quad \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \alpha_i)^2$$

-->

--> $\text{argmin}[\beta[0], \beta[1]] \cdot \text{sum}((y[i] - \alpha[1] \cdot x[i] - \alpha[0])^2, i, 1, n);$

$$(\%o41) \quad \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)^2$$

-->

--> $\text{sum}((y[i] - \alpha[1] \cdot x[i] - \alpha[0])^2, i, 1, n);$

$$(\%o42) \quad \sum_{i=1}^n (y_i - \alpha_1 x_i - \alpha_0)^2$$

--> $\text{diff}(\text{sum}((\text{y}[i] - \alpha[1] \cdot \text{x}[i] - \alpha[0])^2, i, 1, n), \alpha[1], 1);$

$$(\%o44) \quad -2 \sum_{i=1}^n x_i (y_i - \alpha_1 x_i - \alpha_0)$$

--> $-2 \cdot \text{sum}(\text{x}[i] \cdot (\text{y}[i] - \alpha[1] \cdot \text{x}[i] - \alpha[0]), i, 1, n) = 0;$

$$(\%o45) \quad -2 \sum_{i=1}^n x_i (y_i - \alpha_1 x_i - \alpha_0) = 0$$

-->

--> $\text{sum}(\text{x}[i] \cdot (\text{y}[i] - \alpha[1] \cdot \text{x}[i] - \alpha[0]), i, 1, n);$

$$(\%o46) \quad \sum_{i=1}^n x_i (y_i - \alpha_1 x_i - \alpha_0)$$

--> $\text{sum}(\text{x}[i] \cdot \text{y}[i], i, 1, n) - \alpha[1] \cdot \text{sum}(\text{x}[i]^2, i, 1, n) - \alpha[0] \cdot \text{sum}(\text{x}[i], i, 1, n);$

$$(\%o47) \quad \left(\sum_{i=1}^n x_i y_i \right) - \alpha_1 \left(\sum_{i=1}^n x_i^2 \right) - \alpha_0 \sum_{i=1}^n x_i$$

--> $\alpha[0] = \text{sum}(\text{y}[i] - \alpha[1] \cdot \text{x}[i], i, 1, n) / n;$

$$(\%o52) \quad \alpha_0 = \frac{\sum_{i=1}^n y_i - \alpha_1 x_i}{n}$$

--> $\alpha[0] = \text{sum}(\text{y}[i], i, 1, n) / (n) - \alpha[1] \cdot \text{sum}(\text{x}[i], i, 1, n) / n;$

$$(\%o53) \quad \alpha_0 = \frac{\sum_{i=1}^n y_i}{n} - \frac{\alpha_1 \sum_{i=1}^n x_i}{n}$$

--> $\alpha[0] = \text{sum}(\text{y}[i] - \alpha[1] \cdot \text{x}[i], i, 1, n) / n;$

$$(\%o70) \quad \alpha_0 = \frac{\sum_{i=1}^n y_i - \alpha_1 x_i}{n}$$

--> $\alpha[0] = \text{sum}(\text{y}[i] / (n), i, 1, n) - \alpha[1] \cdot \text{sum}(\text{x}[i] / (n), i, 1, n);$

$$(\%o74) \quad \alpha_0 = \frac{\sum_{i=1}^n y_i}{n} - \frac{\alpha_1 \sum_{i=1}^n x_i}{n}$$

--> $(\text{sum}(\text{x}[i] \cdot \text{y}[i], i, 1, n)) - \alpha[1] \cdot (\text{sum}(\text{x}[i]^2, i, 1, n)) - \alpha[0] \cdot \text{sum}(\text{x}[i], i, 1, n);$

$$(\%o55) \quad \left(\sum_{i=1}^n x_i y_i \right) - \alpha_1 \left(\sum_{i=1}^n x_i^2 \right) - \alpha_0 \sum_{i=1}^n x_i$$

$$\begin{aligned}
 & - (\text{sum}(\mathbf{x}[i] \cdot \mathbf{y}[i], i, 1, n)) - \alpha[1] \cdot (\text{sum}(\mathbf{x}[i]^2, i, 1, n)) - (\text{sum}(\mathbf{y}[i], i, 1, n)/n - (\alpha[1] \cdot \text{sum}(\mathbf{x}[i], i, 1, n))/n) \cdot \text{sum}(\mathbf{x}[i], i, 1, n); \\
 & - \\
 & \geq \\
 (\%o75) \quad & \left(\sum_{i=1}^n x_i \right) \left(\frac{\alpha_1 \sum_{i=1}^n x_i}{n} - \frac{\sum_{i=1}^n y_i}{n} \right) + \left(\sum_{i=1}^n x_i y_i \right) - \alpha_1 \sum_{i=1}^n x_i^2
 \end{aligned}$$

$$\text{-->} \quad (\text{sum}(\mathbf{x}[i], i, 1, n)) \cdot (((1/n) \cdot \alpha[1] \cdot \text{sum}(\mathbf{x}[i], i, 1, n)) - (1/n) \cdot \text{sum}(\mathbf{y}[i], i, 1, n)) + (\text{sum}(\mathbf{x}[i] \cdot \mathbf{y}[i], i, 1, n)) - \alpha[1] \cdot \text{sum}(\mathbf{x}[i]^2, i, 1, n);$$

$$(\%o7) \quad \left(\sum_{i=1}^n x_i \right) \left(\frac{\alpha_1 \sum_{i=1}^n x_i}{n} - \frac{\sum_{i=1}^n y_i}{n} \right) + \left(\sum_{i=1}^n x_i y_i \right) - \alpha_1 \sum_{i=1}^n x_i^2$$

--> solve([(%o7)], [\alpha[1]]);

$$(\%o8) \quad [\alpha_1 = \frac{(\sum_{i=1}^n x_i y_i) n - (\sum_{i=1}^n x_i) \sum_{i=1}^n y_i}{(\sum_{i=1}^n x_i^2) n - (\sum_{i=1}^n x_i)^2}]$$

-->

**but wait a minute this was not the result
in the code emporium channel ok let me prove it
to you here is the numerator**

Figure 10:C:\Users\saktheeswaran\Desktop\big.png

$$\arg \min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)^2 \quad \sum_{i=1}^n X_i y_i - \left(\sum_{i=1}^n \frac{y_i}{n} - \widehat{\beta}_1 \sum_{i=1}^n \frac{X_i}{n} \right) \sum_{i=1}^n X_i - \widehat{\beta}_1 \sum_{i=1}^n X_i^2 = 0$$

Differentiating wrt β_1

$$\begin{aligned}
 & \sum_{i=1}^n X_i y_i - \sum_{i=1}^n \frac{y_i}{n} \sum_{i=1}^n X_i + \widehat{\beta}_1 \sum_{i=1}^n \frac{X_i}{n} \sum_{i=1}^n X_i - \widehat{\beta}_1 \sum_{i=1}^n X_i^2 = 0 \\
 & 2 \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)(-X_i) = 0 \quad \sum_{i=1}^n X_i y_i - \frac{1}{n} \sum_{i=1}^n y_i \sum_{i=1}^n X_i + \widehat{\beta}_1 \times \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n X_i - \widehat{\beta}_1 \sum_{i=1}^n X_i^2 = 0 \\
 & \sum_{i=1}^n (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i)(X_i) = 0 \quad \widehat{\beta}_1 = \frac{\sum_{i=1}^n X_i y_i - \frac{1}{n} \sum_{i=1}^n y_i \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n} (\sum_{i=1}^n X_i)^2} \\
 & \sum_{i=1}^n X_i y_i - \widehat{\beta}_0 \sum_{i=1}^n X_i - \widehat{\beta}_1 \sum_{i=1}^n X_i^2 = 0
 \end{aligned}$$



Figure 11:C:\Users\saktheeswaran\Desktop\biger.png

$$\sum_{i=1}^n X_i y_i - \frac{1}{n} \sum_{i=1}^n y_i \sum_{i=1}^n X_i + \widehat{\beta}_1 \times \frac{1}{n} \sum_{i=1}^n X_i \sum_{i=1}^n X_i - \widehat{\beta}_1 \sum_{i=1}^n X_i^2 = 0$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n X_i y_i - \frac{1}{n} \sum_{i=1}^n y_i \sum_{i=1}^n X_i}{\sum_{i=1}^n X_i^2 - \frac{1}{n} (\sum_{i=1}^n X_i)^2}$$

here is the numerator from the code emporium channel

```
--> sum(x[i]*y[i],i,1,n)-(1/n)*(sum(y[i],i,1,n)*sum(x[i],i,1,n));
```

$$(\%o16) \quad \left(\sum_{i=1}^n x_i y_i \right) - \frac{(\sum_{i=1}^n x_i) \sum_{i=1}^n y_i}{n}$$

-->

here is the denominator now i will prove to you the results we got are the same as code emporium channel has derived

```
--> sum(x[i]^2,i,1,n)-(1/n)*(sum(x[i],i,1,n)^2);
```

$$(\%o17) \quad \left(\sum_{i=1}^n x_i^2 \right) - \frac{(\sum_{i=1}^n x_i)^2}{n}$$

-->

divide numerator by denominator you get

```
-- ((sum(x[i]*y[i],i,1,n))-(sum(x[i],i,1,n)*sum(y[i],i,1,n))/n)/((sum(x[i]^2,i,1,n))-(sum(x[i],i,1,n))^2/n)
> ;
```

$$(\%o19) \quad \frac{(\sum_{i=1}^n x_i y_i) - \frac{(\sum_{i=1}^n x_i) \sum_{i=1}^n y_i}{n}}{\left(\sum_{i=1}^n x_i^2\right) - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

now expand it it gets big

--> `expand(%o19);`

$$(\%o20) \quad \frac{\sum_{i=1}^n x_i y_i}{\left(\sum_{i=1}^n x_i^2\right) - \frac{(\sum_{i=1}^n x_i)^2}{n}} - \frac{(\sum_{i=1}^n x_i) \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n x_i^2\right) n - (\sum_{i=1}^n x_i)^2}$$

-->

here is the beautiful part now it gets to the same equation as before

--> `factor(%o20);`

$$(\%o21) \quad \frac{(\sum_{i=1}^n x_i y_i) n - (\sum_{i=1}^n x_i) \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n x_i^2\right) n - (\sum_{i=1}^n x_i)^2}$$

-->

ok did we see this before yes which we got by normal substitution hence are results are correct but we just verified it now go and refer a[1] it will be the same

Figure 12:C:\Users\saktheeswaran\Desktop\fdefd.png



still not finished in this manner the multiple regression can also be derived and lets now go to matrix calculus .org where no computer algebra is capable of dealing with

tensor and matrix calculus now lets see how it can be done

you can also use sympy but for me iam not good at it but i will figure it soon so lets do our matrix calculus there but first lets finish this multiple regression here

Figure 13:C:\Users\saktheeswaran\Desktop\linearregression\11.png

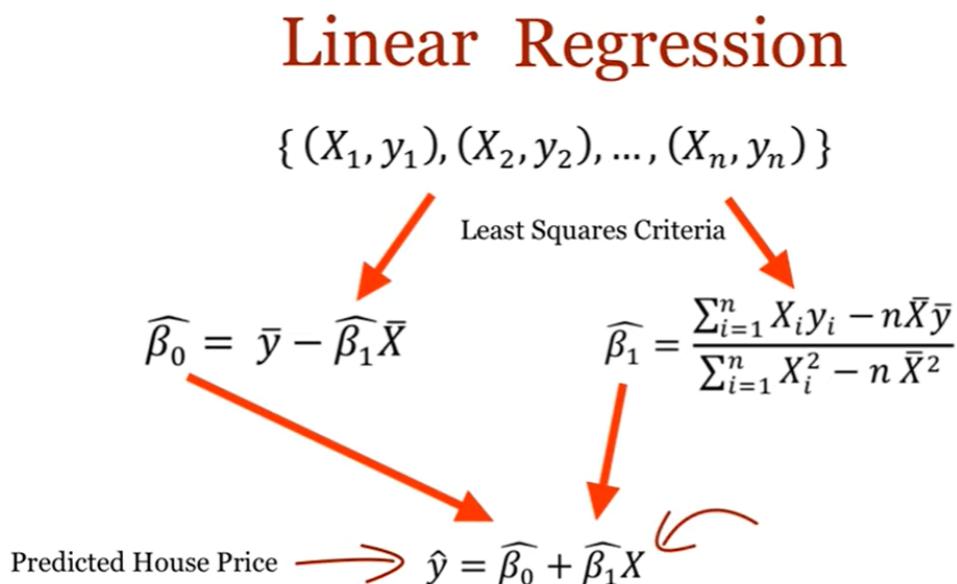


Figure 14:C:\Users\saktheeswaran\Desktop\linearregression\22.png

Multiple Regression

General Parametric Equation:

$$y = \underline{f(\mathbf{X})} + \epsilon$$

Depends on Statistical Method

$$f(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \cdots + \widehat{\beta}_p X_p$$

Figure 15:C:\Users\saktheeswaran\Desktop\linearregression\33.png

Multiple Regression

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \dots & \dots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \dots & \dots & \cdot \\ 1 & X_{3,1} & X_{3,2} & \dots & \dots & \cdot \\ 1 & \cdot & \cdot & \ddots & \ddots & \cdot \\ 1 & \cdot & \cdot & \ddots & \ddots & \cdot \\ 1 & X_{n,1} & X_{n,2} & \dots & \dots & X_{n,p} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

nx1 column matrix

n dimensional column vector or
a matrix nx1

matrix

n dimensional column vector or
a matrix nx1

Figure 16:C:\Users\saktheeswaran\Desktop\linearregression\44.png

Multiple Regression

column vector or nx1 matrix

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \dots & \dots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \dots & \dots & \cdot \\ 1 & X_{3,1} & X_{3,2} & \dots & \dots & \cdot \\ 1 & \cdot & \cdot & \dots & \dots & \cdot \\ 1 & \cdot & \cdot & \dots & \dots & \cdot \\ 1 & X_{n,1} & X_{n,2} & \dots & \dots & X_{n,p} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

eppisilon is an nx1 column vector or nx 1 matrix

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

this is an matrix with nx (p+1) p+1 because multiplicant thats aded
column vector or (p+1)x1 or n dimensional coulmn vector

Figure 17:C:\Users\saktheeswaran\Desktop\linearregression\55.png

Multiple Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{1,2} + \dots + \beta_p X_{1,p} + \epsilon_1 \\ \beta_0 + \beta_1 X_{2,1} + \beta_2 X_{2,2} + \dots + \beta_p X_{2,p} + \epsilon_2 \\ \beta_0 + \beta_1 X_{3,1} + \beta_2 X_{3,2} + \dots + \beta_p X_{3,p} + \epsilon_3 \\ \vdots \\ \vdots \\ \beta_0 + \beta_1 X_{n,1} + \beta_2 X_{n,2} + \dots + \beta_p X_{n,p} + \epsilon_n \end{bmatrix}$$

Figure 18:C:\Users\saktheeswaran\Desktop\linearregression\66.png

Multiple Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_3 - \hat{y}_3 \\ \vdots \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \vdots \\ \hat{y}_n \end{bmatrix} = \mathbf{y} - \hat{\mathbf{y}}$$

Figure 19:C:\Users\saktheeswaran\Desktop\linearregression\77.png

Multiple Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ y_3 - \hat{y}_3 \\ \vdots \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \vdots \\ \hat{y}_n \end{bmatrix} = \mathbf{y} - \hat{\mathbf{y}}$$

$$RSS = \sum_{i=1}^n e_i^2$$

Figure 20:C:\Users\saktheeswaran\Desktop\linearregression\88.png

Multiple Regression

$$RSS = \mathbf{e}^T \mathbf{e}$$

$$RSS = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}})$$

$$RSS = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$RSS = (\mathbf{y}^T - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T)(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$RSS = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}}$$

Figure 21:C:\Users\saktheeswaran\Desktop\linearregression\99.png

Multiple Regression

Matrix Differentiation

$x = m \times 1$ matrix

$A = n \times m$ matrix; $A \perp x$

$$\mathbf{y} = \mathbf{A} \rightarrow \frac{\delta \mathbf{y}}{\delta x} = \mathbf{0}$$

$$\mathbf{y} = \mathbf{Ax} \rightarrow \frac{\delta \mathbf{y}}{\delta x} = \mathbf{A}$$

$$\mathbf{y} = \mathbf{x}\mathbf{A} \rightarrow \frac{\delta \mathbf{y}}{\delta x} = \mathbf{A}^T$$

$$\mathbf{y} = \mathbf{x}^T \mathbf{Ax} \rightarrow \frac{\delta \mathbf{y}}{\delta x} = 2\mathbf{x}^T \mathbf{A}$$

$$RSS = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$\frac{\delta(RSS)}{\delta \hat{\boldsymbol{\beta}}} = \frac{\delta(\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y} + \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}})}{\delta \hat{\boldsymbol{\beta}}} = 0$$

$$\frac{\delta(\mathbf{y}^T \mathbf{y})}{\delta \hat{\boldsymbol{\beta}}} - \frac{\delta(\mathbf{y}^T \mathbf{X}\hat{\boldsymbol{\beta}})}{\delta \hat{\boldsymbol{\beta}}} - \frac{\delta(\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y})}{\delta \hat{\boldsymbol{\beta}}} + \frac{\delta(\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}})}{\delta \hat{\boldsymbol{\beta}}} = 0$$

$$0 - \mathbf{y}^T \mathbf{X} - (\mathbf{X}^T \mathbf{y})^T + 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} = 0$$

$$0 - \mathbf{y}^T \mathbf{X} - \mathbf{y}^T \mathbf{X} + 2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} = 0$$

$$2\hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} = 2\mathbf{y}^T \mathbf{X} \quad \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{X} = \mathbf{y}^T \mathbf{X}$$

$$\hat{\boldsymbol{\beta}}^T = \mathbf{y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Figure 22:C:\Users\saktheeswaran\Desktop\linearregression\100.png

Multiple Regression

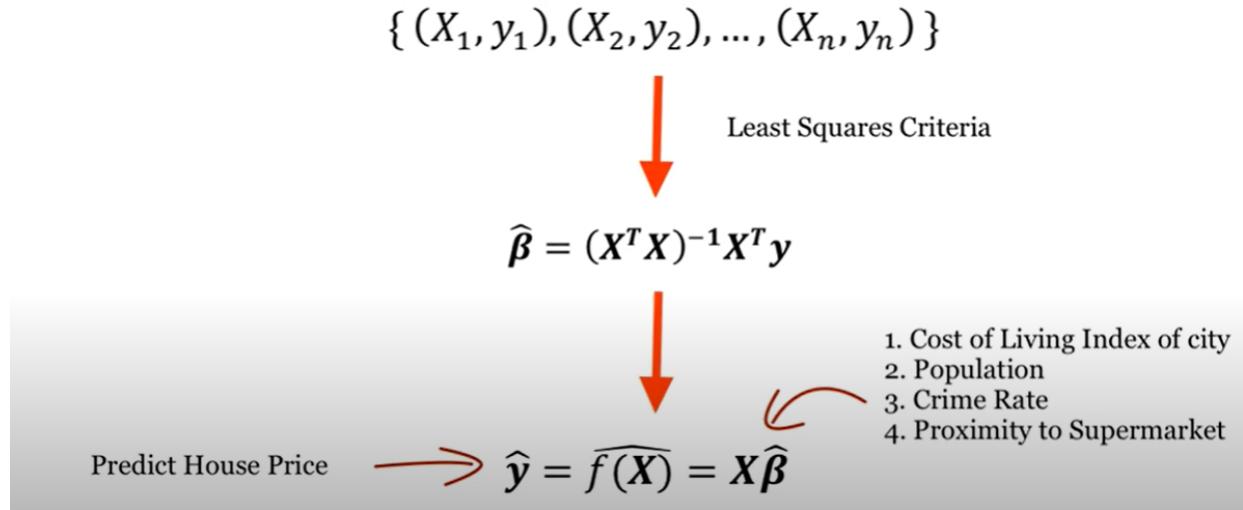


Figure 23:C:\Users\saktheeswaran\Desktop\linearregression\100.png

Multiple Regression

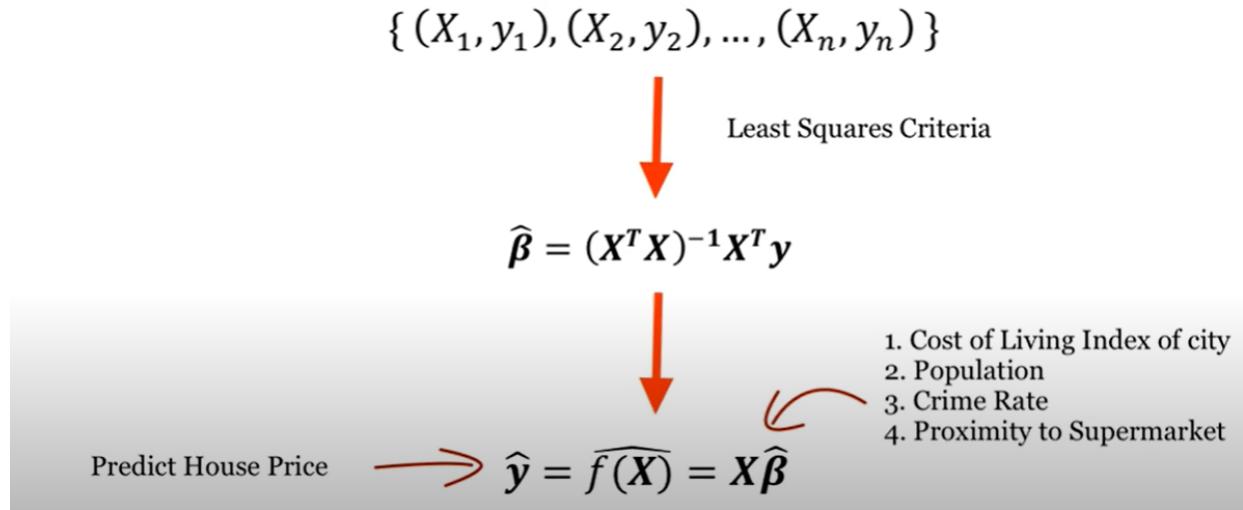


Figure 24:C:\Users\saktheeswaran\Desktop\linearregression\0100u.png

Deriving normal equation for linear regression.

m training examples

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1m} & x_{2m} & \cdots & x_{Nm} \end{bmatrix} \quad W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}$$

$$\text{Error}_i = \begin{cases} y_i - (x_{1i}w_1 + x_{2i}w_2 + \cdots + x_{Ni}w_N) & \rightarrow e_1 \\ y_2 - (x_{12}w_1 + x_{22}w_2 + \cdots + x_{N2}w_N) & \rightarrow e_2 \\ \vdots \\ y_m - (x_{1m}w_1 + x_{2m}w_2 + \cdots + x_{Nm}w_N) & \rightarrow e_m \end{cases} \quad Y - XW$$

f @ techlifewithshivank
© @ techlife_with_shivank

Figure 25:C:\Users\saktheeswaran\Desktop\linearregression\01000v.png

$$\begin{aligned} L &= (Y - XW)^T (Y - XW) \\ &= (Y^T - W^T X^T) (Y - XW) \\ L &= \underline{Y^T Y} - \underline{Y^T X W} - \underline{W^T X^T Y} + \underline{W^T X^T X W} \end{aligned} \quad \left. \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right\}$$

Figure 26:C:\Users\saktheeswaran\Desktop\linearregression\2u.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \mathbf{y}^T \mathbf{x} \omega - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 &\quad + \underline{\nabla_{\omega} (\omega^T \mathbf{x}^T \mathbf{x} \omega)} = \underline{\nabla_{\omega} (\mathbf{x}^T \mathbf{x} \omega)} = \underline{\mathbf{x}^T \mathbf{x} \omega}
 \end{aligned}$$

Figure 27:C:\Users\saktheeswaran\Desktop\linearregression\3v.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \mathbf{y}^T \mathbf{x} \omega - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + \underline{\nabla_{\omega} (\omega^T \mathbf{x}^T \mathbf{x} \omega)} = \underline{\nabla_{\omega} (\mathbf{x}^T \mathbf{x} \omega)} = \underline{\mathbf{x}^T \mathbf{x} \omega}
 \end{aligned}$$

Figure 28:C:\Users\saktheeswaran\Desktop\linearregression\finalgf.png

$$\begin{aligned}
 \Rightarrow \mathbf{x}^T \mathbf{x} \omega &= \mathbf{x}^T \mathbf{y} \\
 \Rightarrow \boxed{\omega = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}} &\quad \text{Normal Eqn}
 \end{aligned}$$

 @techlifewill
 @techlife_w

Figure 29:C:\Users\saktheeswaran\Desktop\linearregression\finalbefore.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \underline{\mathbf{y}^T \mathbf{x} \omega} - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + (\omega^T \mathbf{x}^T \mathbf{x})^T = 0 \\
 &\quad - 2\mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + \mathbf{x}^T \mathbf{x} \omega = 0 \\
 \Rightarrow & 2\mathbf{x}^T \mathbf{x} \omega = 2\mathbf{x}^T \mathbf{y} \\
 \Rightarrow & \mathbf{x}^T \mathbf{x} \omega = \mathbf{x}^T \mathbf{y} \\
 \Rightarrow & \omega = \mathbf{x}^T
 \end{aligned}$$

@techlifewithshivan
 @techlife_with_shiva

Figure 30:C:\Users\saktheeswaran\Desktop\linearregression\extraaa22.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{x} \omega - \omega^T \mathbf{x}^T \mathbf{y} + \omega^T \mathbf{x}^T \mathbf{x} \omega
 \end{aligned}
 \quad \left. \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right\}$$

$$\nabla_{\omega} L = \begin{pmatrix} \frac{\partial L}{\partial \omega_1} \\ \frac{\partial L}{\partial \omega_2} \\ \vdots \\ \frac{\partial L}{\partial \omega_N} \end{pmatrix} = 0 \quad \begin{pmatrix} \omega_1 \\ \omega_2 \\ | \\ \omega_N \end{pmatrix}$$

@techlifewithshivan
 @techlife_with_shiva

Figure 31:C:\Users\saktheeswaran\Desktop\linearregression\etragghexplain11.png

$$\begin{aligned}
 L &= (\gamma - x\omega)^T (\gamma - x\omega) \\
 &= (\gamma^T - \omega^T x^T) (\gamma - x\omega) \\
 L &= \gamma^T \gamma - \gamma^T x\omega - \omega^T x^T \gamma + \omega^T x^T x\omega
 \end{aligned}
 \quad \left| \quad \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right.$$

$$\nabla_{\omega} L =$$

$$\nabla_{\omega}(\omega^T a) = \nabla_{\omega}(a^T \omega) = a$$

$$\omega^T a = a^T \omega = \omega_1 a_1 + \omega_2 a_2 + \dots + \omega_n a_n$$

$$\begin{pmatrix} \frac{\partial(\omega^T a)}{\partial \omega_1} \\ | \\ \frac{\partial(\omega^T a)}{\partial \omega_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ | \\ a_n \end{pmatrix} = a$$

 @ techlifewithshivank
 @ techlife_with_shivank

Figure 32:C:\Users\saktheeswaran\Desktop\linearregression\extra23344.png

$$\begin{aligned}
 L &= (\gamma - x\omega)^T (\gamma - x\omega) \\
 &= (\gamma^T - \omega^T x^T) (\gamma - x\omega) \\
 L &= \gamma^T \gamma - \gamma^T x\omega - \omega^T x^T \gamma + \omega^T x^T x\omega
 \end{aligned}
 \quad \left| \quad \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right.$$

$$\nabla_{\omega} L =$$

$$\nabla_{\omega}(\omega^T x) = \nabla_{\omega}(x^T \omega)$$

 @ techlifewithshivank
 @ techlife_with_shivank

Figure 33:C:\Users\saktheeswaran\Desktop\linearregression\extraaa22.png

$$\begin{aligned}
 L &= (\gamma - x\omega)^T (\gamma - x\omega) \\
 &= (\gamma^T - \omega^T x^T) (\gamma - x\omega) \\
 L &= \gamma^T \gamma - \gamma^T x\omega - \omega^T x^T \gamma + \omega^T x^T x\omega
 \end{aligned}
 \quad \left. \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right\}$$

$$\nabla_{\omega} L = \begin{pmatrix} \frac{\partial L}{\partial \gamma} \\ \frac{\partial L}{\partial \omega_1} \\ \frac{\partial L}{\partial \omega_2} \\ \vdots \\ \frac{\partial L}{\partial \omega_N} \end{pmatrix} = 0 \quad \begin{pmatrix} \omega_1 \\ \omega_2 \\ | \\ \omega_N \end{pmatrix}$$

f @techlifewithshiva
o @techlife_with_shiva

Figure 34:C:\Users\saktheeswaran\Desktop\linearregression\1u.png

$$\begin{aligned}
 L &= (\gamma - x\omega)^T (\gamma - x\omega) \\
 &= (\gamma^T - \omega^T x^T) (\gamma - x\omega) \\
 L &= \underbrace{\gamma^T \gamma} - \underbrace{\gamma^T x\omega} - \underbrace{\omega^T x^T \gamma} + \underbrace{\omega^T x^T x\omega}_{(A+B)^T = A^T + B^T} \\
 \nabla_{\omega} L &= 0 - x^T \gamma - x^T \gamma + (A\omega)^T = A^T \omega + \omega^T A
 \end{aligned}
 \quad \begin{array}{l} (A\omega)^T = \omega^T A \\ (AB)^T = B^T A^T \end{array}$$

$$d(uv) = d(u)v + u d(v)$$

Figure 35:C:\Users\saktheeswaran\Desktop\linearregression\2u.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \mathbf{y}^T \mathbf{x} \omega - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega}
 \end{aligned}$$

$$\begin{aligned}
 (A+B)^T &= A^T + B^T \\
 (AB)^T &= B^T A^T
 \end{aligned}$$

$$\nabla_{\omega} (\underline{\omega^T \mathbf{a}}) = \nabla_{\omega} (\underline{\mathbf{a}^T \omega}) = \underline{\mathbf{a}}$$

Figure 36:C:\Users\saktheeswaran\Desktop\linearregression\3v.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \mathbf{y}^T \mathbf{x} \omega - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega}
 \end{aligned}$$

$$\begin{aligned}
 (A+B)^T &= A^T + B^T \\
 (AB)^T &= B^T A^T
 \end{aligned}$$

$$\nabla_{\omega} (\underline{\omega^T \mathbf{a}}) = \nabla_{\omega} (\underline{\mathbf{a}^T \omega}) = \underline{\mathbf{a}}$$

Figure 37:C:\Users\saktheeswaran\Desktop\linearregression\4v.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \underline{\mathbf{y}^T \mathbf{x} \omega} - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + (\omega^T \mathbf{x}^T \mathbf{x})^T
 \end{aligned}$$

$$\begin{aligned}
 (A+B)^T &= A^T + B^T \\
 (AB)^T &= B^T A^T
 \end{aligned}$$

$$\nabla_{\omega} (\underline{\omega^T \mathbf{a}}) = \underline{\nabla_{\omega} (\mathbf{a}^T \omega)} = \underline{\mathbf{a}}$$

Figure 38:C:\Users\saktheeswaran\Desktop\linearregression\5v.png

$$\begin{aligned}
 L &= (\mathbf{y} - \mathbf{x}\omega)^T (\mathbf{y} - \mathbf{x}\omega) \\
 &= (\mathbf{y}^T - \omega^T \mathbf{x}^T) (\mathbf{y} - \mathbf{x}\omega) \\
 L &= \underline{\mathbf{y}^T \mathbf{y}} - \underline{\mathbf{y}^T \mathbf{x} \omega} - \underline{\omega^T \mathbf{x}^T \mathbf{y}} + \underline{\omega^T \mathbf{x}^T \mathbf{x} \omega} \\
 \nabla_{\omega} L &= 0 - \mathbf{x}^T \mathbf{y} - \mathbf{x}^T \mathbf{y} + \mathbf{x}^T \mathbf{x} \omega + (\omega^T \mathbf{x}^T \mathbf{x})^T = 0 \\
 &\quad - \cancel{\omega^T \mathbf{x}^T \mathbf{y}} + \mathbf{x}^T \mathbf{x} \omega + \cancel{\mathbf{x}^T \mathbf{x} \omega} = 0
 \end{aligned}$$

$$\begin{aligned}
 (A+B)^T &= A^T + B^T \\
 (AB)^T &= B^T A^T \\
 \uparrow &
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow \quad 2 \mathbf{x}^T \mathbf{x} \omega &= 2 \mathbf{x}^T \mathbf{y} \\
 \Rightarrow \quad \mathbf{x}^T \mathbf{x} \omega &= \mathbf{x}^T \mathbf{y} \\
 \Rightarrow
 \end{aligned}$$

Figure 39:C:\Users\saktheeswaran\Desktop\linearregression\6vvv.png

$$\begin{aligned} \partial X^T x \omega &= \partial x^T y \\ x^T x \omega &= x^T y \\ \boxed{\omega = (x^T x)^{-1} x^T y} \end{aligned}$$

Figure 40:C:\Users\saktheeswaran\Desktop\linearregression\extraaa22.png

$$\begin{aligned} L &= (y - x\omega)^T (y - x\omega) \\ &= (y^T - \omega^T x^T) (y - x\omega) \\ L &= y^T y - y^T x \omega - \omega^T x^T y + \omega^T x^T x \omega \end{aligned} \quad \left| \begin{array}{l} (A+B)^T = A^T + B^T \\ (AB)^T = B^T A^T \end{array} \right.$$

$$\nabla_{\omega} L = \begin{pmatrix} \frac{\partial L}{\partial \omega_1} \\ \frac{\partial L}{\partial \omega_2} \\ \vdots \\ \frac{\partial L}{\partial \omega_N} \end{pmatrix} = 0 \quad \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{pmatrix}$$

 @ techlifewithshiva
 @ techlife_with_shiva

Figure 41:C:\Users\saktheeswaran\Desktop\matrix-calculus.png

The screenshot shows a web browser window for 'Matrix Calculus' at matrixcalculus.org. The page title is 'Matrix Calculus'. A sidebar on the left says 'Sponsored by data assessment solutions. The AI-powered platform for Skills and Resource Management www.decidalo.com'. The main content area has a form for calculating derivatives. The input field contains 'derivative of $y^T \cdot y - y^T \cdot x \cdot b - b^T \cdot x \cdot y + c \cdot b^T + d \cdot b$ ' w.r.t. 'b'. Below the form, the derivative is shown as:

$$\frac{\partial}{\partial b} (y^T \cdot y - y^T \cdot x \cdot b - b^T \cdot x \cdot y + c \cdot b^T + d \cdot b) = c \cdot \mathbb{T} - ((y^T \cdot x)^T \otimes \mathbb{I} + \mathbb{I} \otimes (x^T \cdot y)) + d \cdot \mathbb{I} \otimes \mathbb{I}$$

where

Export functions as Python Latex
 Common subexpressions ON

Input fields for variable types:
 b is a matrix
 c is a scalar
 d is a scalar
 x is a matrix
 y is a matrix

Figure 42:C:\Users\saktheeswaran\Desktop\matrix-calculusd.png

The screenshot shows a web browser window for 'Matrix Calculus' at matrixcalculus.org. The page title is 'Matrix Calculus'. A sidebar on the left says 'Sponsored by data assessment solutions. The AI-powered platform for Skills and Resource Management www.decidalo.com'. The main content area has a form for calculating derivatives. The input field contains 'derivative of $y^T \cdot y - y^T \cdot x \cdot b - b^T \cdot x \cdot y + c \cdot b^T + d \cdot b$ ' w.r.t. 'b'. Below the form, the derivative is shown as:

$$\frac{\partial}{\partial b} (y^T \cdot y - y^T \cdot x \cdot b - b^T \cdot x \cdot y + c \cdot b^T + d \cdot b) = c \cdot \mathbb{T} - ((y^T \cdot x)^T \otimes \mathbb{I} + \mathbb{I} \otimes (x^T \cdot y)) + d \cdot \mathbb{I} \otimes \mathbb{I}$$

where

Export functions as Python Latex
 Common subexpressions ON

Input fields for variable types:
 b is a matrix
 c is a scalar
 d is a scalar
 x is a matrix
 y is a matrix