Multiple Linear Regression

In Simple Linear Regression →

$$Y = a + bX$$
$$Y = \theta_0 + \theta_1 x$$ → 2 equations to solve

For 2 input features →

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$ → 3 eqn to solve

For d parameters (Input features) →

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_d x_d$$ → (d+1) eqn to solve

It will be a tedious job to solve this eqn, so we will make a <u>compact representaion</u> to solve it comparitively easier

* For large number of features (d), solving d+1 equations is a tedious approach

* This necessitates a compact representation leading to a well-defined soln.

$$\hat{Y} = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \cdots + \theta_d X_d$$

X_1	X_2	X_3	X_4	$\cdots$	X_d	Y
X_{11}	X_{12}	X_{13}	X_{14}	$\cdots$	X_{1d}	Y_1
X_{21}	X_{22}	X_{23}	X_{24}	$\cdots$	X_{2d}	Y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{n1}	X_{n2}	X_{n3}	X_{n4}	$\cdots$	X_{nd}	Y_n

$\longleftrightarrow$ d input features

Again we have to minimize $\sum_{i=1}^{n} \mathcal{E}_i^2$ where, n are # of datapoints

- Error of datapoint 1 $(E_1) = Y_1 - (\theta_0 + \theta_1 X_{11} + \theta_2 X_{12} + \cdots + \theta_d X_{1d})$
- Error of datapoint 2 $(E_2) = Y_2 - (\theta_0 + \theta_1 X_{21} + \theta_2 X_{22} + \cdots + \theta_d X_{2d})$

$\vdots$

- Error of datapoint n $(E_n) = Y_n - (\theta_0 + \theta_1 X_{n1} + \theta_2 X_{n2} + \cdots + \theta_d X_{nd})$

$$\underbrace{\begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ \vdots \\ E_n \end{bmatrix}}_{\substack{\text{Error Matrix} \\ (E_{n\times 1})}} = \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}}_{\substack{\text{Actual Value Matrix} \\ (Y_{n\times 1})}} - \underbrace{\begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1d} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2d} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3d} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nd} \end{bmatrix}}_{\text{Coefficient Matrix } (X_{n\times(d+1)})} \times \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}}_{\substack{\text{Parameter Matrix} \\ (\theta_{(d+1)\times 1})}}$$

$$E_{n\times 1} = Y_{n\times 1} - X_{n\times(d+1)} \cdot \theta_{(d+1)\times 1}$$

$$\Rightarrow \boxed{E_{n\times 1} = Y_{n\times 1} - (X\theta)_{n\times 1}}$$

* We have to minimize $\sum_{i=1}^{n} \varepsilon_i^2$

$$\sum_{i=1}^{n} E_i^2 = E_1 E_1 + E_2 E_2 + E_3 E_3 + \cdots + E_n E_n$$

$$= \begin{bmatrix} E_1 & E_2 & E_3 & \cdots & E_n \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ \vdots \\ E_n \end{bmatrix}$$

$$\sum_{i=1}^{n} E_i^2 = E^T . E = (Y - X\theta)^T (Y - X\theta)$$

$$= (Y - X\theta)^T (Y - X\theta)$$

$$= (Y^T - \theta^T X^T)(Y - X\theta)$$

$$= Y^T Y - Y^T X\theta - \theta^T X^T Y + \theta^T X^T X\theta$$

<u>To Minimize :-</u>

$$\nabla (E^T . E)_\theta = \frac{\partial (E^T E)}{\partial \theta} = 0$$

$$= \cancel{\frac{\partial}{\partial \theta}(Y^T Y)}^{\,0} - \frac{\partial}{\partial \theta}(Y^T X\theta) - \frac{\partial}{\partial \theta}(\theta^T X^T Y) + \frac{\partial}{\partial \theta}(\theta^T X^T X\theta)$$

$$\frac{\partial}{\partial\theta}(Y^TX\theta) \Rightarrow$$ We know, if $f(\theta) = \alpha^T\theta + \beta$,
then $\nabla f(\theta) = (\alpha^T)^T$

Hence, if $f(\theta) = Y^TX\theta$,
then $\nabla f(\theta) = (Y^TX)^T = X^TY$

$$\frac{\partial}{\partial\theta}(\theta^TX^TY) \Rightarrow$$ Here $\theta^TX^TY = (X^TY)^T\theta$
$= Y^TX\theta$

$f(\theta) = Y^TX\theta$
$\nabla f(\theta) = (Y^TX)^T = X^TY$

$$\frac{\partial}{\partial\theta}(\theta^TX^TX\theta) \Rightarrow$$ We know, $f(\theta) = \theta^TA\theta$
then, $\nabla f(\theta) = (\theta^T + \theta)\omega$

Hence, if $f(\theta) = \theta^TX^TX\theta$
then, $\nabla f(\theta) = (X^TX + X^TX)\theta$

Putting values in $\nabla(E^T.E)_\theta$

$$\nabla(E^TE)_\theta = 0 - X^TY - X^TY + 2X^TX\theta = 0$$
$$= -2X^TY + 2X^TX\theta = 0$$
$$2(-X^TY + X^TX\theta) = 0$$
$$-X^TY + X^TX\theta = 0$$
$$X^TX\theta = X^TY$$

$$\boxed{\theta = (X^TX)^{-1}(X^TY)}$$

Note ÷ if $Y = \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_d X_d$

then Cofficient matrix $X = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nd} \end{bmatrix}_{n \times d}$

$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}_{d \times 1}$$

* Computation Cost →

$\dim(X) = n \times (d+1)$

$\dim(X^T) = (d+1) \times n$

$\dim(Y) = n \times 1$

$$\theta = (X^T X)^{-1} X^T Y$$

$(d+1)\times n$ $\quad n\times(d+1)$ $\quad (d+1)\times n$ $\quad n\times 1$

$O(nd^2)$ $\quad O(nd)$

$[\;]_{p\times q} \; [\;]_{q\times r}$

Total multiplication = pqr

Solving a $d \times d$ system of eqn costs $O(d^3)$: Using Guassian Elimination

Total Cost = $O(nd^2 + d^3)$

Consider the example below where the mass, y (grams), of a chemical is related to the time, x(seconds), for which the chemical reaction has been taking place according to the table:

Time, x (seconds)	Mass, y (grams)
5	40
7	120
12	180
16	210
20	240

Can be done with Simple Linear Regression →

$$\overline{Y} = a + b\overline{X}$$
$$\overline{XY} = a\overline{X} + b\overline{X^2}$$

	X	Y	XY	X^2
1	5	40	200	25
2	7	120	840	49
3	12	180	2160	144
4	16	210	3360	256
5	20	240	4800	400
	60	790	11360	874
	12	158	2272	174.8

$$158 = a + 12b$$
$$2272 = 12a + 174.8b$$

$$\boxed{\begin{aligned} a &= -11.5065 \\ b &= -12.2078 \end{aligned}}$$

Estimate the values of ω_1 and ω_2 by minimizing the loss function $L(\theta)$, where $\theta = [\omega_1, \omega_2]$. The loss function is defined as the mean squared error between the predicted and actual values of y:

$$\theta^* = \arg\min_\theta L(\theta)$$
$$L(\theta) = \frac{1}{N}\sum_{n=1}^{N}(f_\theta(x_n) - y_n)^2$$

where $f_\theta(x_n) = \omega_1 x_{1,n} + \omega_2 x_{2,n}$ and N is the number of data points.

Data Table:

x_1	x_2	y
-3.5	1.9	0.21
2.1	-3.3	0.01
-4.3	5.1	0.97
1.7	2.2	0.98
0.1	-0.1	0.45

$$X = \begin{bmatrix} -3.5 & 1.9 \\ 2.1 & -3.3 \\ -4.3 & 5.1 \\ 1.7 & 2.2 \\ 0.1 & -0.1 \end{bmatrix} \qquad Y = \begin{bmatrix} 0.21 \\ 0.01 \\ 0.97 \\ 0.98 \\ 0.45 \end{bmatrix}$$

$$\theta = (X^T X)^{-1} X^T Y$$

$$(X^T X) = \underset{2\times5}{\begin{bmatrix} -3.5 & 2.1 & -4.3 & 1.7 & 0.1 \\ 1.9 & -3.3 & 5.1 & 2.2 & -0.1 \end{bmatrix}} \underset{5\times2}{\begin{bmatrix} -3.5 & 1.9 \\ 2.1 & -3.3 \\ -4.3 & 5.1 \\ 1.7 & 2.2 \\ 0.1 & -0.1 \end{bmatrix}}$$

$$(X^T X) = \begin{bmatrix} 38.05 & -31.78 \\ -31.78 & 45.36 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{(38.05 \times 45.36) - (-31.78 \times -31.78)} \begin{bmatrix} 45.36 & 31.78 \\ 31.78 & 38.05 \end{bmatrix}$$

$$= \frac{1}{715.98} \begin{bmatrix} 45.36 & 31.78 \\ 31.78 & 38.05 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.0633 & 0.0444 \\ 0.0444 & 0.0531 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \underset{2\times2}{\begin{bmatrix} 0.0633 & 0.0444 \\ 0.0444 & 0.0531 \end{bmatrix}} \underset{2\times5}{\begin{bmatrix} -3.5 & 2.1 & -4.3 & 1.7 & 0.1 \\ 1.9 & -3.3 & 5.1 & 2.2 & -0.1 \end{bmatrix}}$$

$$= \begin{bmatrix} -0.13719 & -0.01359 & -0.04575 & 0.20529 & 0.00189 \\ -0.05451 & -0.08199 & 0.07989 & 0.1923 & -0.00087 \end{bmatrix}$$

$$(X^T X)^{-1} X^T Y =$$

$$\underset{2 \times 5}{\begin{bmatrix} -0.13719 & -0.01359 & -0.04575 & 0.20529 & 0.00189 \\ -0.05451 & -0.08199 & 0.07989 & 0.1923 & -0.00087 \end{bmatrix}} \underset{5 \times 1}{\begin{bmatrix} 0.21 \\ 0.01 \\ 0.97 \\ 0.98 \\ 0.45 \end{bmatrix}}$$

$$\theta = \begin{bmatrix} 0.1287114 \\ 0.253288 \end{bmatrix}$$

$$\boxed{f_\theta(x_n) = 0.1287114\, x_{1,n} + 0.253288\, x_{2,n}}$$

Question: Given a training dataset of 500 instances, with each input instance having 6 dimensions (features) and each output being a scalar value, we wish to apply **linear regression** to this data. The linear regression model is represented by the equation:

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_6 x_6$$

What are the dimensions of the design matrix used in applying linear regression to this data?

Options:

(A) 500×6

(B) 500×7

(C) 500×6^2

(D) None of the above

$$\begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,6} \\ X_2 & X_{2,2} & \cdots & X_{2,6} \\ \vdots & \vdots & \ddots & \vdots \\ X_{500,1} & X_{500,2} & \cdots & X_{500,6} \end{bmatrix}$$

(A) 500×6 Ans.

Problem:
You are training a linear regression model on two disjoint datasets, **A** and **B**, where $|A| \gg |B|$. The model minimizes the following sum of squared errors across both datasets:

$$L_{\text{standard}}(w) = \sum_{i=1}^{|A|} \left(w^T x_a^{(i)} - y_a^{(i)}\right)^2 + \sum_{i=1}^{|B|} \left(w^T x_b^{(i)} - y_b^{(i)}\right)^2$$

How will the model most likely treat the two groups in terms of performance?

Options:

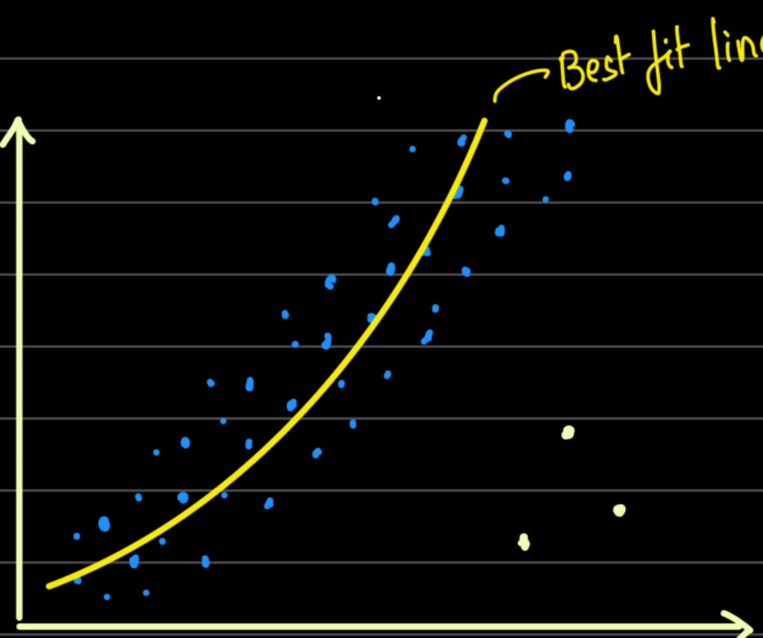
1. The model will give more importance to group **A**, leading to better performance on **A** and worse performance on **B**.
2. The model will give more importance to group **B**, leading to better performance on **B** and worse performance on **A**.
3. The model will treat both groups **A** and **B** equally, without bias toward either group.

We have 1003 datapoints $\rightarrow$ $|A| = 1000$, $|B| = 3$

$$L = \underbrace{\left(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \cdots + \varepsilon_{1000}^2\right)}_{\Downarrow \text{ For dataset A}} + \underbrace{\left(\varepsilon_{1001}^2 + \varepsilon_{1002}^2 + \varepsilon_{1003}^2\right)}_{\Downarrow \text{ For dataset B}}$$

(More influence on dataset A)

(1) Ans.



Problem:
You are training a linear regression model on two disjoint datasets, **A** and **B**, where $|A| \gg |B|$. The current loss function is:

$$L_{\text{standard}}(w) = \sum_{i=1}^{|A|} \left(w^T x_a^{(i)} - y_a^{(i)} \right)^2 + \sum_{i=1}^{|B|} \left(w^T x_b^{(i)} - y_b^{(i)} \right)^2$$

Which of the following loss functions should you use to ensure more equal treatment of both datasets **A** and **B**?

Options:

1. Use the same loss function as above (no change).
2. Divide by $|A|$ for dataset **A** and by $|B|$ for dataset **B**.
3. Divide dataset **A**'s loss by $|B|$ and dataset **B**'s loss by $|A|$.

$$L: \underbrace{\left(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \cdots + \varepsilon_{1000}^2\right)}_{\Downarrow \text{ For dataset A}} + \underbrace{\left(\varepsilon_{1001}^2 + \varepsilon_{1002}^2 + \varepsilon_{1003}^2\right)}_{\Downarrow \text{ For dataset B}}$$

To increase influence on B, we can take value α, β where $|\beta| \gg |\alpha|$ such that:

$$L = \alpha\left(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \cdots + \varepsilon_{1000}^2\right) + \beta\left(\varepsilon_{1001}^2 + \varepsilon_{1002}^2 + \varepsilon_{1003}^2\right)$$

We know $|A| \gg |B|$, i.e. $\frac{1}{|B|} \gg \frac{1}{|A|}$

$$L = \frac{1}{|A|}\left(\varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_{1000}^2\right) + \frac{1}{|B|}\left(\varepsilon_{1001}^2 + \varepsilon_{1002}^2 + \varepsilon_{1003}^2\right)$$

(2) Ans.

Problem:

You want to estimate a quantity y as a function of x. Suppose you decide to model your estimate $\hat{y}$ as follows:

$$\hat{y} = w_0 + w_1 \log(x) \Rightarrow \quad y = w_0 + w_1 t$$

use normal equation

Given the data:

y	x
5	1
6	2
7	3
8	4

Find the values of w_0 and w_1 that minimize the mean squared error (MSE) for the given data (up to 2 decimal places).

$$\hat{y} = w_0 + w_1 \log(x)$$

$$MSE = \frac{1}{N} \sum_{1}^{N} (y - w_0 - w_1 \log(x))^2$$

let $t = \log(x)$

x	$\log(x) = t$	y	t^2	$t \cdot y$
1	0	5	0	0
2	0.69314	6	0.48044	4.15884
3	1.09861	7	1.20694	7.69027
4	1.38629	8	1.92179	11.09032

$\overline{t} = 0.79451$ $\qquad \overline{t^2} = 0.90229$

$\overline{y} = 6.5$ $\qquad \overline{t \cdot y} = 5.73485$

$$\overline{y} = w_0 + w_1 \overline{t} \qquad 6.5 = w_0 + 0.79451 w_1$$

$$\overline{t \cdot y} = w_0 \overline{t} + w_1 \overline{t^2} \qquad 5.73485 = 0.79451 w_0 + 0.90229 w_1$$

Solving System of Equation gives →

$$w_0 = 4.82 \qquad w_1 = 2.10495$$

$$\boxed{\hat{y} = 4.82 + 2.10495 \log(x)}$$

* What if X or $(X^T X)$ is not invertible??

We know the closed form solution of Multiple LR is given by:

$$\theta = (X^T X)^{-1} X^T Y$$

But if X is not invertible $(X^T X)^{-1}$ is also not invertible hence closed form solution for such dataset does not exist.

To overcome this problem, we either apply regularisation (which will be discussed in Ridge and Lasso Regression).

OR

We use Moore-Penrose closed-form solution:

$$\theta = X^+ y$$

where X^+ is the pseudoinverse of X (the coefficient matrix), which is computed using SVD {Singular Value Decomposition}

SVD of X is given as:

$$X = U \Sigma V^T$$

where, $U \in R^{n \times n}$ is an orthogonal matrix of left singular vectors of X
$V \in R^{p \times p}$ is an orthogonal matrix of right singular vectors of X.
$\Sigma \in R^{n \times p}$ contains singular values.

Pseudoinverse X^+ is computed as:

$$X^+ = V \Sigma^+ U^T$$

where Σ^+ is formed by taking the reciprocal of each non-zero singular value in Σ.

Thus, the closed-form solution for θ using Moore-Penrose pseudoinverse is :

$$\boxed{\theta = V\Sigma^+ U^T y}$$

So, when does inverse of X or X^TX fail to exist?

1. More features than data point :

   If $X \in R^{n \times p}$ and $p > n$, then X^TX is guaranteed to be rank deficient (not invertible)

$$\text{Rank}(X^TX) = \text{Rank}(X)$$

2. Linear dependence among columns of X :

   If some columns of coefficient matrix X are linearly dependent, then

$$\text{Rank}(X) < p \qquad \text{where} \quad X \in R^{n \times p}$$

   Hence, inverse of X will not exist.

Example : Given dataset →

X_1	X_2	Y
1	2	1
2	4	2
3	6	3

$$\text{Coefficient matrix } X = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}_{3\times 2}$$

- Here column 2 of X = 2* Column 1 of X {linearly dependent}
- $\text{Rank}(X^TX) = \text{Rank}(X) = 1$
- Hence, inverse of X^TX fails to exist.

Applying Moore-Penrose pseudoinverse closed-form solution :

We know SVD of any matrix $X = U\Sigma V^T$

$$\begin{aligned} X^TX &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U \Sigma V^T \qquad \because U^TU = I \\ &= V\Sigma^T\Sigma V^T \end{aligned}$$

$$X^TX = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 14 & 28 \\ 28 & 56 \end{bmatrix}$$

Eigen values of $X^TX \Rightarrow \begin{vmatrix} 14-\lambda & 28 \\ 28 & 56-\lambda \end{vmatrix} = 0$

$$\lambda^2 - 70\lambda = 0$$
$$\lambda(\lambda - 70) = 0$$
$$\lambda = 0, 70$$

Eigen vectors corresponding $\lambda = 70$:

$$\begin{bmatrix} 14-70 & 28 \\ 28 & 56-70 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -56 & 28 \\ 28 & -14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_2 \rightarrow R_2 + \frac{1}{2} R_1$$

$$\begin{bmatrix} -56 & 28 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Let $x_2 = K$

$$-56x_1 + 28K = 0$$
$$x_1 = \frac{28}{56} K$$
$$x_1 = \frac{1}{2} K$$

$$v_1 = \frac{1}{\sqrt{\left(\frac{1}{2}K\right)^2 + K^2}} \begin{bmatrix} 1/2\, K \\ K \end{bmatrix}$$

$$= \frac{1}{\frac{K}{2}\sqrt{5}} K \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$$

$$= \frac{2}{\sqrt{5}} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}$$

$$v_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Eigen vectors corresponding $\lambda = 0$:

$$\begin{bmatrix} 14 & 28 \\ 28 & 56 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_2 \rightarrow R_2 - 2R_1$$

$$\begin{bmatrix} 14 & 28 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Let $x_2 = K$

$$14x_1 + 28K = 0$$

$$x_1 = \frac{-28}{14} K$$

$$x_1 = -2K$$

$$v_2 = \frac{1}{\sqrt{K^2 + (-2K)^2}} \begin{bmatrix} -2K \\ K \end{bmatrix}$$

$$= \frac{1}{K\sqrt{5}} K \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$$v_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

$$V = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{70} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$XX^T = U\Sigma V^T (U \Sigma V^T)^T$$
$$= U\Sigma V^T V \Sigma^T U^T \qquad \because V^T V = I$$
$$= U \Sigma \Sigma^T U^T$$

$$XX^T = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 10 & 15 \\ 10 & 20 & 30 \\ 15 & 30 & 45 \end{bmatrix}$$

Eigen values of $XX^T \Rightarrow \begin{vmatrix} 5-\lambda & 10 & 15 \\ 10 & 20-\lambda & 30 \\ 15 & 30 & 45-\lambda \end{vmatrix} = 0$

$$\lambda^3 - 70\lambda^2 = 0$$
$$\lambda^2(\lambda - 70) = 0$$
$$\lambda = 0, 0, 70$$

Eigen vectors corresponding $\lambda = 70$:

$$\begin{bmatrix} 5-70 & 10 & 15 \\ 10 & 20-70 & 30 \\ 15 & 30 & 45-70 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -65 & 10 & 15 \\ 10 & -50 & 30 \\ 15 & 30 & -25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$R_2 \rightarrow R_2 + \frac{10}{65} R_1$$

$$R_3 \rightarrow R_3 + \frac{15}{65} R_1$$

$$\begin{bmatrix} -65 & 10 & 15 \\ 0 & -630/13 & 420/13 \\ 0 & 420/13 & -280/13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$R_3 \rightarrow R_3 + \frac{2}{3} R_2$$

$$\begin{bmatrix} -65 & 10 & 15 \\ 0 & -630/13 & 420/13 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$-65x_1 + 10x_2 + 15x_3 = 0$$
$$-630x_2 + 420x_3 = 0$$

Let $x_3 = k$:

$$-630x_2 = -420k$$
$$x_2 = \frac{2}{3}k$$

$$-65x_1 + \frac{20}{3}k + 15k = 0$$
$$x_1 = \frac{1}{3}k$$

$$u_1 = \frac{1}{\sqrt{\left(\frac{1}{3}k\right)^2 + \left(\frac{2}{3}k\right)^2 + k^2}} \begin{bmatrix} 1/3\,k \\ 2/3\,k \\ k \end{bmatrix}$$

$$= \frac{3}{k\sqrt{14}}\, k \begin{bmatrix} 1/3 \\ 2/3 \\ 1 \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 1/\sqrt{14} \\ 2/\sqrt{14} \\ 3/\sqrt{14} \end{bmatrix}$$

Eigen vectors corresponding $\lambda = 0$:

$$\begin{bmatrix} 5 & 10 & 15 \\ 10 & 20 & 30 \\ 15 & 30 & 45 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$R_2 \to R_2 - 2R_1$$
$$R_3 \to R_3 - 3R_1$$

$$\begin{bmatrix} 5 & 10 & 15 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$5x_1 + 10x_2 + 15x_3 = 0$$

1. Let $x_2 = K, \; x_3 = 0$:

$$5x_1 + 10K = 0$$
$$x_1 = -2K$$

$$u_2 = \frac{1}{\sqrt{(-2K)^2 + K^2}} \begin{bmatrix} -2K \\ K \\ 0 \end{bmatrix}$$

$$= \frac{1}{K\sqrt{5}} K \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

$$u_2 = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}$$

2. Let $x_2 = 0 \quad x_3 = K$:

$$5x_1 + 15K = 0$$
$$x_1 = -3K$$

$$U_3 = \frac{1}{\sqrt{(-3K)^2 + K^2}} \begin{bmatrix} -3K \\ 0 \\ K \end{bmatrix}$$

$$= \frac{1}{K\sqrt{10}} K \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

$$U_3 = \begin{bmatrix} -3/\sqrt{10} \\ 0 \\ 1/\sqrt{10} \end{bmatrix}$$

$$U = \begin{bmatrix} 1/\sqrt{14} & -2/\sqrt{5} & -3/\sqrt{10} \\ 2/\sqrt{14} & 1/\sqrt{5} & 0 \\ 3/\sqrt{14} & 0 & 1/\sqrt{10} \end{bmatrix}$$

Now, we have our complete Singular Value Decomposition of X :

$$X = \underbrace{\begin{bmatrix} 1/\sqrt{14} & -2/\sqrt{5} & -3/\sqrt{10} \\ 2/\sqrt{14} & 1/\sqrt{5} & 0 \\ 3/\sqrt{14} & 0 & 1/\sqrt{10} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{70} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}}_{V^T}$$

Computing Moore-penrose pseudoinverse :

$$X^+ = V\Sigma^+ U^T$$

$$X^+ = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 1/\sqrt{70} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{14} & 2/\sqrt{14} & 3/\sqrt{14} \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 \\ -3/\sqrt{10} & 0 & 1/\sqrt{10} \end{bmatrix}$$

$$X^+ = \begin{bmatrix} 1/\sqrt{350} & 0 & 0 \\ 2/\sqrt{350} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{14} & 2/\sqrt{14} & 3/\sqrt{14} \\ -2/\sqrt{5} & 1/\sqrt{5} & 0 \\ -3/\sqrt{10} & 0 & 1/\sqrt{10} \end{bmatrix}$$

$$X^+ = \begin{bmatrix} 1/70 & 2/70 & 3/70 \\ 2/70 & 4/70 & 6/70 \end{bmatrix}$$

Computing $\theta = X^+ y$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1/70 & 2/70 & 3/70 \\ 2/70 & 4/70 & 6/70 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 2/5 \end{bmatrix}$$

$$\boxed{\hat{y} = 0.2\, x_1 + 0.4\, x_2}$$