Linear Regression

Anubha Gupta, PhD.
Professor
SBILab, Dept. of ECE,
IIIT-Delhi, India
Contact: anubha@iiitd.ac.in; Lab: http://sbilab.iiitd.edu.in

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

Linear Regression

Learning Objectives

In this module, we will learn to

- solve a regression problem
  - Using the statistical approach
  - Using the data Driven approach

- Regression errors and performance metrics
- Example

Linear Regression
(using the statistical approach)

Multiple Linear Regression

- A linear regression technique that uses several (more than two) explanatory variables to predict the outcome of a response variable.

n observations

- **System of linear equations**

$$\left\{\begin{aligned} y_1 &= \theta_0 + \theta_1 x_{1,1} + \theta_2 x_{1,2} + \cdots + \theta_{p-1} x_{1,p-1} + \varepsilon_1 \\ y_2 &= \theta_0 + \theta_1 x_{2,1} + \theta_2 x_{2,2} + \cdots + \theta_{p-1} x_{2,p-1} + \varepsilon_2 \\ &\vdots \\ y_n &= \theta_0 + \theta_1 x_{n,1} + \theta_2 x_{n,2} + \cdots + \theta_{p-1} x_{n,p-1} + \varepsilon_n \end{aligned}\right.$$

- **Matrix Notation**

$$[\mathbf{y}]_{n\mathrm{x}1} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,(p-1)} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,(p-1)} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,(p-1)} \end{bmatrix}_{n\times p} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{p-1} \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

Data matrix

$$\underline{y} = X\underline{\theta} + \underline{\epsilon} \qquad \hat{\underline{y}}$$

Maximum Likelihood Estimation (MLE)

$\underline{x}_1^T$

$$[\mathbf{y}]_{n\text{x}1} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,(p-1)} \\ 1 & x_{2,1} & x_{2,2} & & x_{2,(p-1)} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,(p-1)} \end{bmatrix}_{n\times p} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{p-1} \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

$y_1, \underline{x}_1$

$y_2, \underline{x}_2$

$\vdots$

$y_n, \underline{x}_n$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \vdots \\ \mathbf{x}_{\mathrm{n}}^{\mathrm{T}} \end{pmatrix} \boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\boxed{\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}}$$

Maximum Likelihood Estimation (MLE)

$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

Random variable $\boxed{y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + \varepsilon_i} \rightarrow$ r.v

$$E[y_i] = E[\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + \varepsilon_i] = E[\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}] + E[\varepsilon_i]$$

$$m_{y_i} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta} + 0 = x_i^T\theta = \langle x_i, \theta \rangle$$

$$E[(y_i - m_{y_i})^2] = \sigma_{y_i}^2$$

$$= E\left[\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2\right] = E[\varepsilon_i^2]$$

$$E[(\varepsilon_i)^2] = \sigma^2$$

$$y_i \sim \mathcal{N}(x_i^T\theta, \sigma^2)$$

Maximum Likelihood estimator

$$L'(\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2}{2\sigma^2}}$$

[Using the assumptions of linear regression]

$$L'(\boldsymbol{\theta}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}_i, \boldsymbol{\theta})$$

[Maximize the Likelihood function]

$$\text{MLE: } \hat{\theta} = \max_{\boldsymbol{\theta}} \; p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$$

$$\max_{\boldsymbol{\theta}} \; L'(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta}} \; p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$$

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \; exp\left(-\sum_{i=1}^{n} \frac{\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2}{2\sigma^2}\right)$$

Maximum Likelihood estimator

$$L(\boldsymbol{\theta}) = \log \ p(\mathbf{y}|\cancel{\mathbf{X}}, \boldsymbol{\theta}) = - \ \log\left((2\pi)^{\frac{n}{2}}\sigma^n\right) - \sum_{i=1}^{n} \frac{(y_i - \mathbf{x}_i^T\boldsymbol{\theta})^2}{2\sigma^2}$$ [Log likelihood function]

$$\boxed{\log_e e^a = a}$$

Maximum Likelihood estimator:

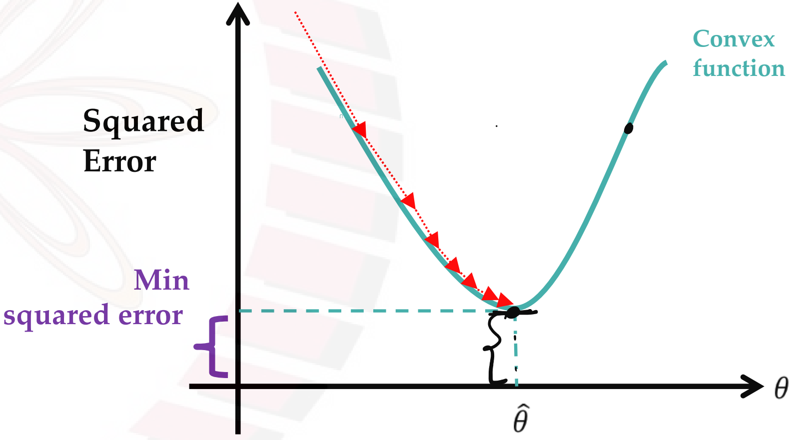
M~~inimization of~~ Max. $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$:

$$\underset{\theta}{\text{Minimization of}} \left(\sum_{i=1}^{n} (\underbrace{y_i}_{y_i} - \underbrace{\mathbf{x}_i^T\boldsymbol{\theta}}_{\hat{y}_i})^2\right) = \min_{\theta}\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right] = \min_{\theta}\left(\sum_{i=1}^{n} \varepsilon_i^2\right)$$

Maximum Likelihood estimator

MLE leads to Least Squares Solutions

$$\hat{\theta}_{mle} = \min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2$$



Linear Regression Solution

objective fn. $\rightarrow$ $\boxed{L(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T\boldsymbol{\theta})^2}$

Let us consider a case with two features:

parameters (unknown)

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \varepsilon_i$$

error of i^{th} sample

i^{th} sample

features of i^{th} sample

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}$$

Linear Regression proof

$$L(\underline{\theta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$y_i = \underbrace{\theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2}}_{\hat{y}_i} + \varepsilon_i$$

$$L(\boldsymbol{\theta}) = (y_1 - \underbrace{\theta_0 - x_{11}\theta_1 - x_{12}\theta_2}_{\hat{y}_1})^2 + (y_2 - \underbrace{\theta_0 - x_{21}\theta_1 - x_{22}\theta_2}_{\hat{y}_2})^2 + \ldots.. + (y_n - \underbrace{\theta_0 - x_{n1}\theta_1 - x_{n2}\theta_2}_{\hat{y}_n})^2$$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_0} = 2(y_1 - \theta_0 - x_{11}\theta_1 - x_{12}\theta_2)(\underline{-1}) + 2(y_2 - \theta_0 - x_{21}\theta_1 - x_{22}\theta_2)(-1) + \ldots.. + 2(y_n - \theta_0 - x_{n1}\theta_1 - x_{n2}\theta_2)(-1)$$

$$\boxed{\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_0} = 2\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta})\,(-1)}$$

Linear Regression proof

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_1} = 2(y_1 - \theta_0 - x_{11}\theta_1 - x_{12}\theta_2)(-x_{11}) + 2(y_2 - \theta_0 - x_{21}\theta_1 - x_{22}\theta_2)(-x_{21})$$
$$+ \;\ldots.. \;+ 2(y_n - \theta_0 - x_{n1}\theta_1 - x_{n2}\theta_2)(-x_{n1})$$

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_1} = 2\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta})\,(-x_{i1})$$
$$\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_2} = 2\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta})\,(-x_{i2})$$

$-\underline{x}_i$

$$\underline{\nabla L(\boldsymbol{\theta})} = \begin{bmatrix} \frac{\partial L(\theta)}{\partial \theta_0} \\ \frac{\partial L(\theta)}{\partial \theta_1} \\ \frac{\partial L(\theta)}{\partial \theta_2} \end{bmatrix} = 2\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}) \begin{bmatrix} -1 \\ -x_{i1} \\ -x_{i2} \end{bmatrix} = 0$$

$$\boxed{-2\sum_{i=1}^{n}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta})\mathbf{x}_i = 0}$$

Closed form solution

$$-2\sum_{i=1}^{n}\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)\mathbf{x}_i = 0 \Rightarrow \sum_{i=1}^{n} y_i \underline{x}_i - \sum_{i=1}^{n} \underline{x}_i(\underline{x}_i^T \underline{\theta}) = 0$$

$\langle \underline{x}_i, \underline{\theta} \rangle$

$$\sum_{i=1}^{n} y_i\mathbf{x}_i = \sum_{i=1}^{n} \mathbf{x}_i\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)$$

$$X = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix}$$

$$y_1\mathbf{x}_1 + y_2\mathbf{x}_2 + \cdots + y_n\mathbf{x}_n = \mathbf{x}_1\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\theta}\right) + \mathbf{x}_2\left(\mathbf{x}_2^{\mathrm{T}}\boldsymbol{\theta}\right) + \cdots + \mathbf{x}_n\left(\mathbf{x}_n^{\mathrm{T}}\boldsymbol{\theta}\right)$$

$$\underbrace{\begin{bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_n \end{bmatrix}}_{X^T} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \underbrace{\begin{bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_n \end{bmatrix}}_{X^T} \underbrace{\begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix}}_{X} \underline{\theta}$$

$$\boxed{\mathbf{X}^{\mathrm{T}}\mathbf{y} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\boldsymbol{\theta}}$$

$$\boxed{\boldsymbol{\theta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}} = \quad \underline{\theta} = X^{\dagger}\underline{y}$$

$$y = X\underline{\theta} + \underline{\varepsilon}$$

Closed form solution

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \boldsymbol{\theta})^2 \Rightarrow \min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \Rightarrow \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 = \min_{\theta} \|\epsilon\|_2^2$$

pseudo inverse/Moore Penrose inverse

$[\mathbf{X}]_{nxp}$ = tall matrix

$n > p$

$$\mathbf{X}^{\dagger} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

$$\boldsymbol{\theta} = \mathbf{X}^{\dagger}\mathbf{y}$$

If $\mathbf{X}$ is a square invertible matrix

$$\mathbf{X}^{\dagger} = \mathbf{X}^{-1}(\mathbf{X}^T)^{-1}\mathbf{X}^T$$

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

$$= \mathbf{X}^{-1}$$

$$\boldsymbol{\theta} = \mathbf{X}^{\dagger}\mathbf{y}$$

$$\boldsymbol{\theta} = \mathbf{X}^{-1}\mathbf{y}$$

$$\boxed{\mathbf{y} = \mathbf{X}\boldsymbol{\theta}}$$

Geometric Interpretation: MLE

$$\underline{\mathbf{X}} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{bmatrix} 1 & \cdots & x_{\cancel{p-1},1,p-1} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{\cancel{p-1},n,p-1} \end{bmatrix} = [\mathbf{z}_1\ \mathbf{z}_2\ \mathbf{z}_3\ \ldots\ \mathbf{z}_p]$$

data matrix

$\langle z_1, \epsilon \rangle = 0 = z_1^T \epsilon = 0$

$\langle z_2, \epsilon \rangle = 0 = z_2^T \epsilon = 0$

$$\begin{bmatrix} z_1^T \\ \vdots \\ z_p^T \end{bmatrix} \epsilon = 0$$

$$\mathbf{X}^T \boldsymbol{\varepsilon} = 0$$

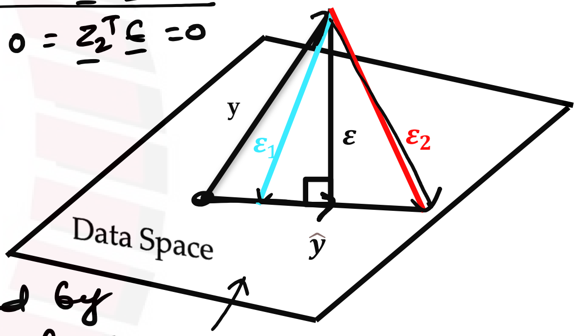
$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = 0$$

$$\mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}\boldsymbol{\theta} = 0$$

$$(\mathbf{X}^T\mathbf{X})\boldsymbol{\theta} = \mathbf{X}^T\mathbf{y}$$

$$\boxed{\boldsymbol{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{P}\mathbf{y}}$$

$\underline{y} = X\underline{\theta} + \underline{\epsilon}$



spanned by the columns of X

P is the projection matrix which ~~is~~ does the orthogonal projection of y on the data space.

Drawbacks of MLE

1. **Sensitivity to outliers**: MLE assumes a specific probability distribution for the data which can produce biased estimates if the data contains outliers.
2. **Lack of robustness**: Small changes in the data can lead to large changes in the estimates of the parameters. This can be a problem if the data is noisy or contains errors.
3. **Overfitting**: MLE can suffer from overfitting if the model is too complex or if the data is limited. Overfitting occurs when the model fits the noise in the data rather than the underlying pattern, which can lead to poor generalization performance on new data.
4. **Computational complexity**: MLE can be computationally complex, especially for high-dimensional data or complex models. This can make it difficult to estimate the parameters or to perform inference on the model.

NPTEL

Linear Regression
(Data Driven Approach)

Linear Regression Solution

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2$$

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}\right)^2 \Longrightarrow \min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \Longrightarrow \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \Rightarrow \min_{\underline{\theta}} \|\epsilon\|_2^2$$

$$\boldsymbol{\theta} = \mathbf{X}^{\dagger}\mathbf{y}$$

where

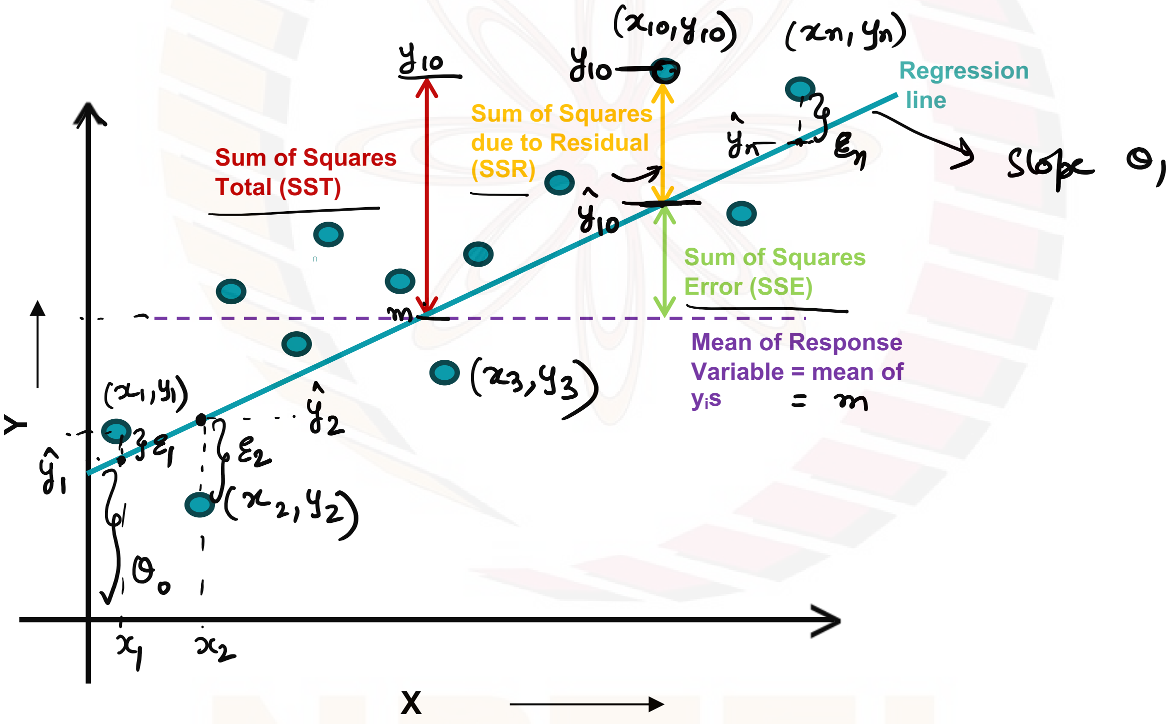
$$\boxed{\mathbf{X}^{\dagger} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}}$$

Linear Regression
(Types of Errors and Performance Metrics)

Types of Errors and Performance Metrics

- Regression errors:
  - Sum of Squares Total (SST) / Total Squared Error (TSS)
  - Sum of Squares due to Regression (SSR)
  - Sum of Squares error (SSE)

- Performance Metrics:
  - R-squared score
  - Mean Absolute error
  - Mean Squared error
  - Root Mean Squared error


Sum of Squares Total (SST)
Sum of Squares due to Residual (SSR)
Sum of Squares Error (SSE)
Regression line
Mean of Response Variable = mean of y_is = m
Slope θ_1
θ_0
(x_{10}, y_{10})
(x_n, y_n)
(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
y_{10}
$\hat{y}_{10}$
$\hat{y}_n$
$\hat{y}_1$
$\hat{y}_2$
ε_1
ε_2
ε_n
m
x_1
x_2
X
Y

Regression Errors

- Sum of Squares Total (SST) / Total Squared Error (TSS)

$$SST = \sum_{i=1}^{n} (y_i - m)^2$$

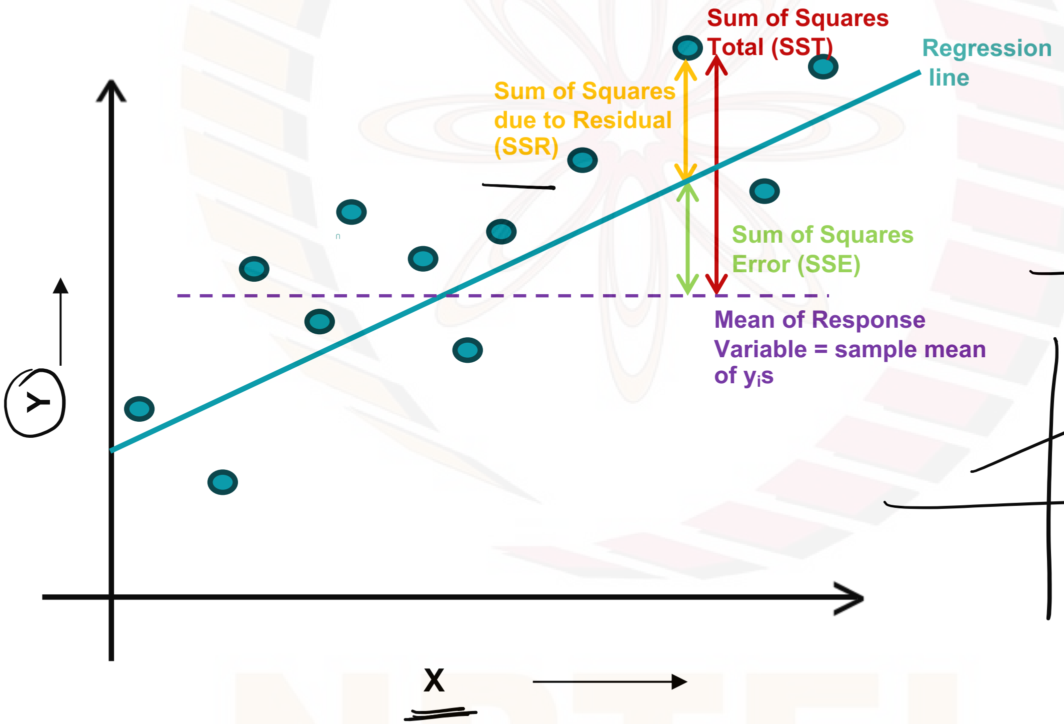
- Sum of Squares due to Regression (SSR)/ Residual Sum of Squares (RSS)

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

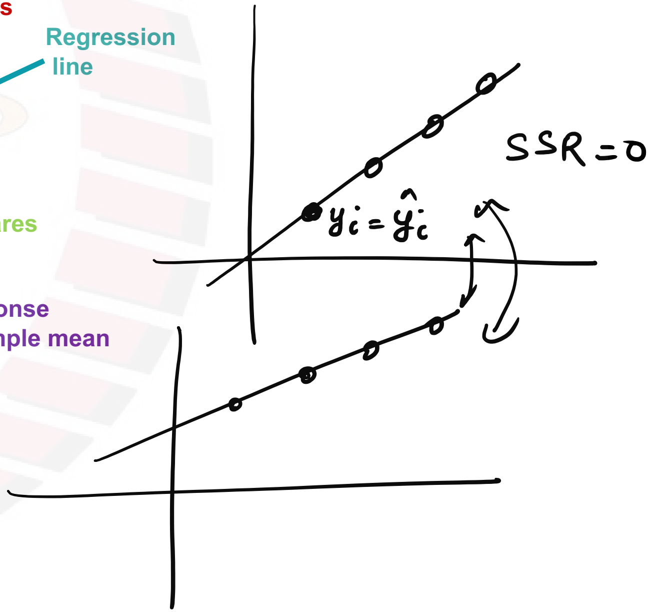
- Sum of Squares error (SSE) / Explained Sum of Squares (ESS)

$$SSE = \sum_{i=1}^{n} (m - \hat{y}_i)^2$$

$m = \sum_{i=1}^{n} y_i$: mean value of the various actual values



Sum of Squares Total (SST)
Regression line
Sum of Squares due to Residual (SSR)
Sum of Squares Error (SSE)
Mean of Response Variable = sample mean of y_is
Y
X



$y_i = \hat{y}_i$
SSR = 0

Performance Metric: R-Squared (R^2) Error

- Also known as the **Coefficient of Determination**, it is given by

$$\mathrm{R}^2 = 1 - \left(\frac{SSR}{SST}\right) = \frac{SSE}{SST}$$

- The measure of the proportion of the variance in the dependent variable that can be explained by the independent variables in the model.

    - $R^2 = 1 \Longrightarrow$ the model explains all the variance in the dependent variable
    - $R^2 = 0 \Longrightarrow$ the model does not explain any of the variance in the dependent variable

- Measures the goodness of fit or best-fit-line

Performance Metric: Mean Absolute Error

- The measure of average absolute difference between the predicted and actual values of a set of observations

ε_1 ε_2

$$\boxed{MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}$$

where,

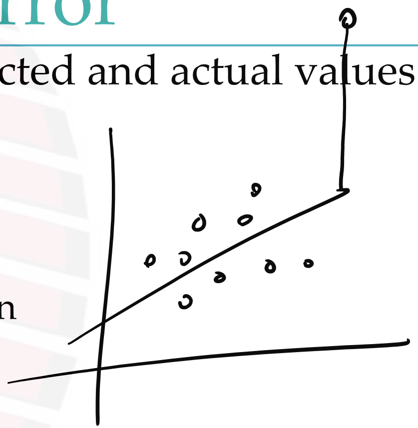
y_i: actual value of the i^{th} observation
$\hat{y}_i$: predicted value of the i^{th} observation
n : total number of observations

- A lower MAE value indicates better performance of the model because it implies that the predictions are closer to the actual values
- **Advantages**:
  - Gives an idea of how far off the predictions are from the actual values, in the same units as the original data
  - Less sensitive to outliers than some other metrics, such as Mean Squared Error (MSE)
  - Treats all errors (+ve or –ve) equally
- **Disadvantages**:
  - Gives equal weight to all errors, regardless of their values

Performance Metric: Mean Squared Error

- The measure of average squared difference between the predicted and actual values of a set of observations

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



where,

y_i: actual value of the ith observation
$\hat{y}_i$: predicted value of the ith observation
n : total number of observations

- **Advantages**:
  - Gives an idea of how far off the predictions are from the actual values
  - Penalizes large errors more heavily than small errors

- **Disadvantages**:
  - Sensitive to outliers (a few large errors can have a disproportionate impact on the metric)

Performance Metric: Root Mean Squared Error

- The measure of square root of the average squared difference between the predicted and actual values of a set of observations

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where,

y_i: actual value of the i^{th} observation
$\hat{y}_i$: predicted value of the i^{th} observation
n : total number of observations

- **Advantages**:
    - Unlike MSE, its range is suppressed due to the square root and the magnitude of error is comparatively lower.

- **Disadvantages**:
    - Similar to MSE

Let us Try!

- Regression errors:
  - Sum of Squares Total (SST):
  - Sum of Squares due to Regression (SSR):
  - Sum of Squares error (SSE):

Explanatory variable (feature) x	Response variable (output) y
1 x_1	2 y_1
2 x_2	3 y_2
3 x_3	5 y_3
4 x_4	6 y_4
5 x_5	8 y_5

mean m

$n = 5$

Performance Metric
- R-squared score:
- Mean Absolute error:
- Mean Squared error:
- Root Mean Squared error:

$y_i \,,\ \hat{y}_i$

$$\boxed{\hat{y}_i = \theta_0 + \theta_1 x_i}$$

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} \qquad \underline{\theta} = X^{\dagger} \underline{y} \qquad \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \qquad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

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

Let us Try!

- Regression errors:
  - Sum of Squares Total (SST): 22.8
  - Sum of Squares due to Regression (SSR): 0.3
  - Sum of Squares error (SSE): 22.5

x

Performance Metric
- R-squared score: 0.9869
- Mean Absolute error: 0.240
- Mean Squared error: 0.060
- Root Mean Squared error: 0.245

Explanatory variable (feature)	Response variable (output)
1	2
2	3
3	5
4	6
5	8

NPTEL

Try it Yourself!

$\Rightarrow n=50$

Question: Let there be a dataset of 50 observations, where each data point consists of a single independent variable (x) and a single dependent variable (y). You want to fit a linear regression model to this data to predict y based on x. After performing the regression analysis, you obtain the following regression equation: $\boxed{\hat{y} = 2.5 + 1.8x}$

Using the above information, what would be the predicted value of y for a value of x equal to 3.2?

The solution to this question will not be provided.

To Summarize

In this module, we discussed on

- What is linear regression?
- What are the different types of errors?
- How to solve the problem?
- We also looked at one example.