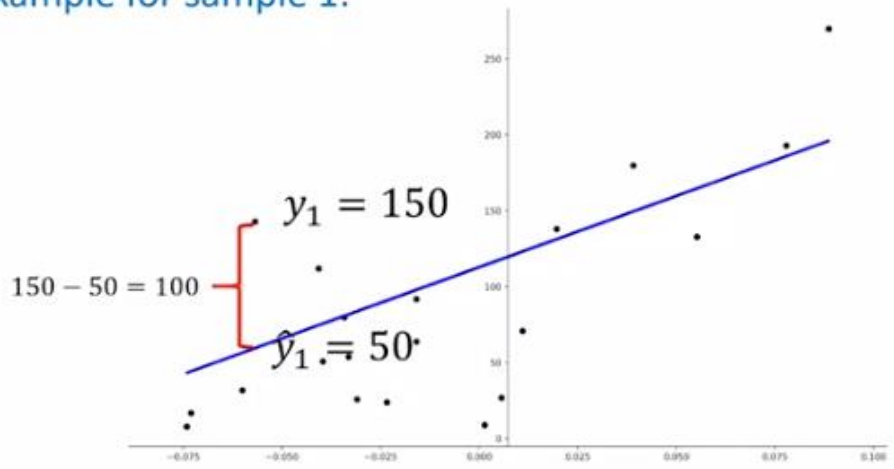
Measures for In-Sample Evaluation

- A way to numerically determine how good the model fits on dataset.

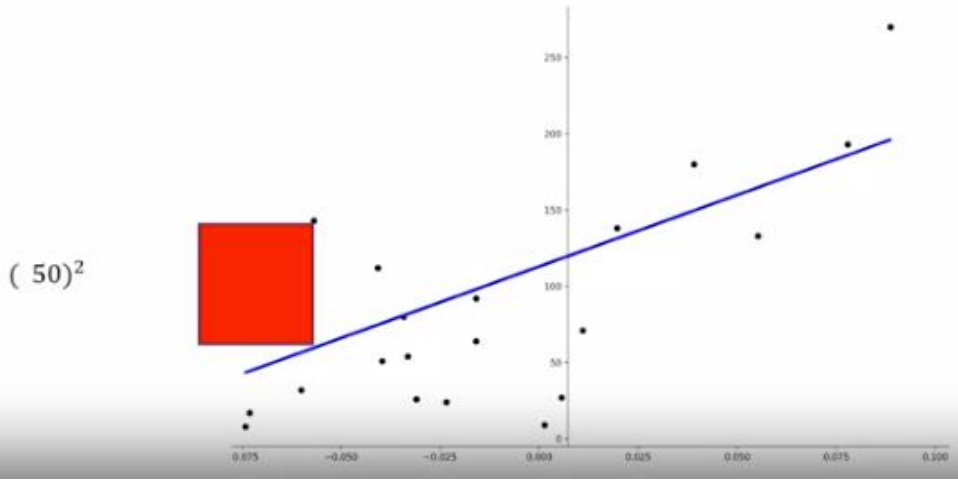
- Two important measures to determine the fit of a model:
  - **Mean Squared Error (MSE)**
  - **R-squared (R^2)**

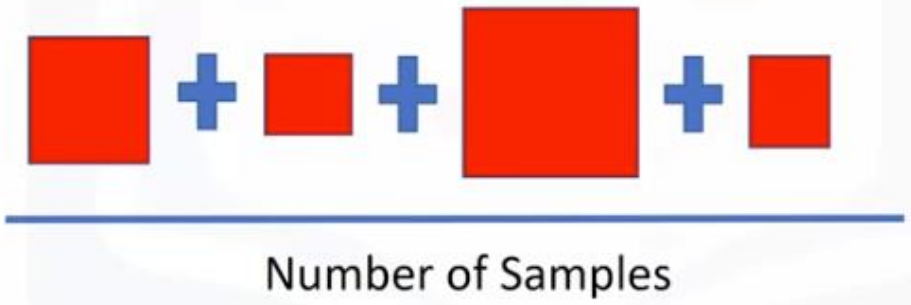
Mean Squared Error (MSE)

- For Example for sample 1:



- To make all the values positive we square it



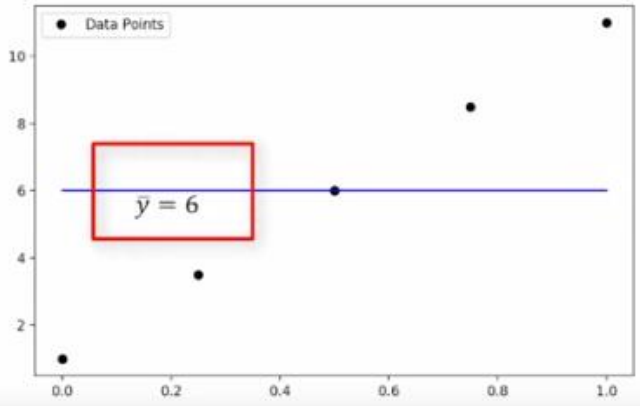


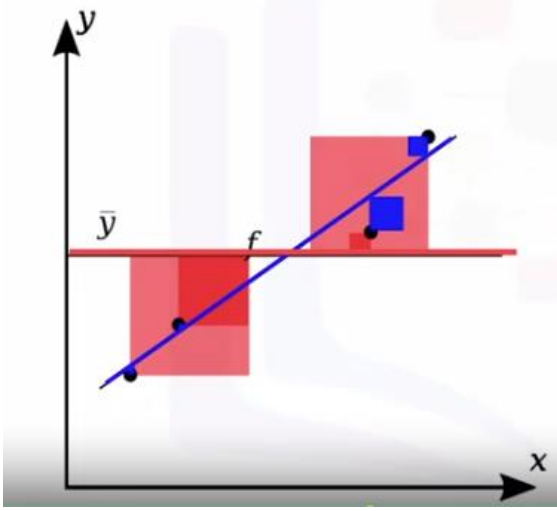
R-squared/ R^2

- The Coefficient of Determination or R squared (R^2)
- Is a measure to determine how close the data is to the fitted regression line.
- R^2: the percentage of variation of the target variable (Y) that is explained by the linear model.
- Think about as comparing a regression model to a simple model i.e the mean of the data points

$$R^2 = \left(1 - \frac{\text{MSE of regression line}}{\text{MSE of the average of the data}}\right)$$

- In this example the average of the data points $\bar{y}$ is 6

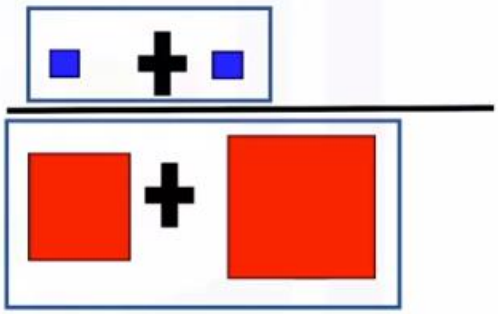


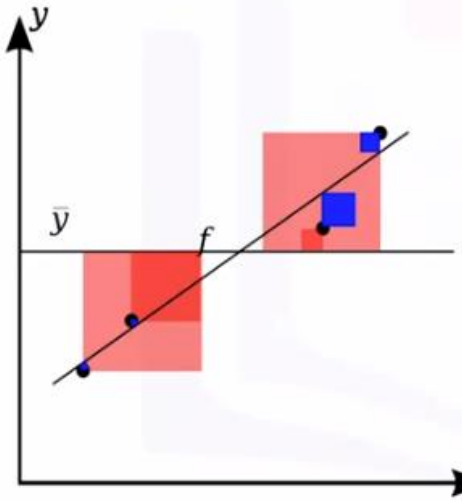


- The blue line represents the regression line
- The blue squares represents the MSE of the regression line
- The red line represents the average value of the data points
- The red squares represent the MSE of the red line
- We see the area of the blue squares is much smaller than the area of the red squares

- In this case ratio of the areas of MSE is close to zero

$$\frac{\text{MSE of regression line}}{\text{MSE of } \bar{y}} = 0$$



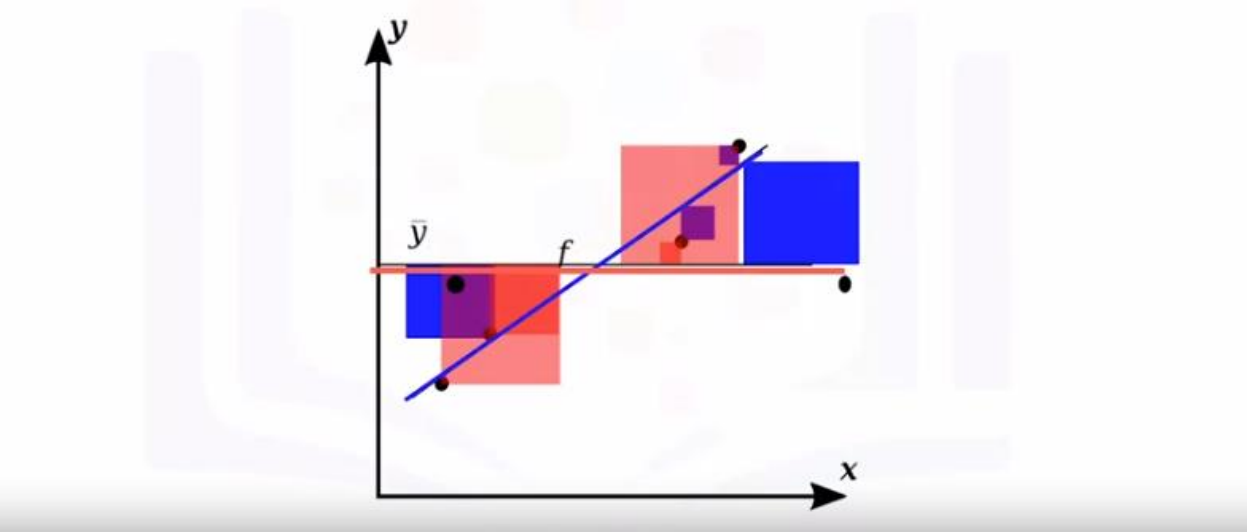


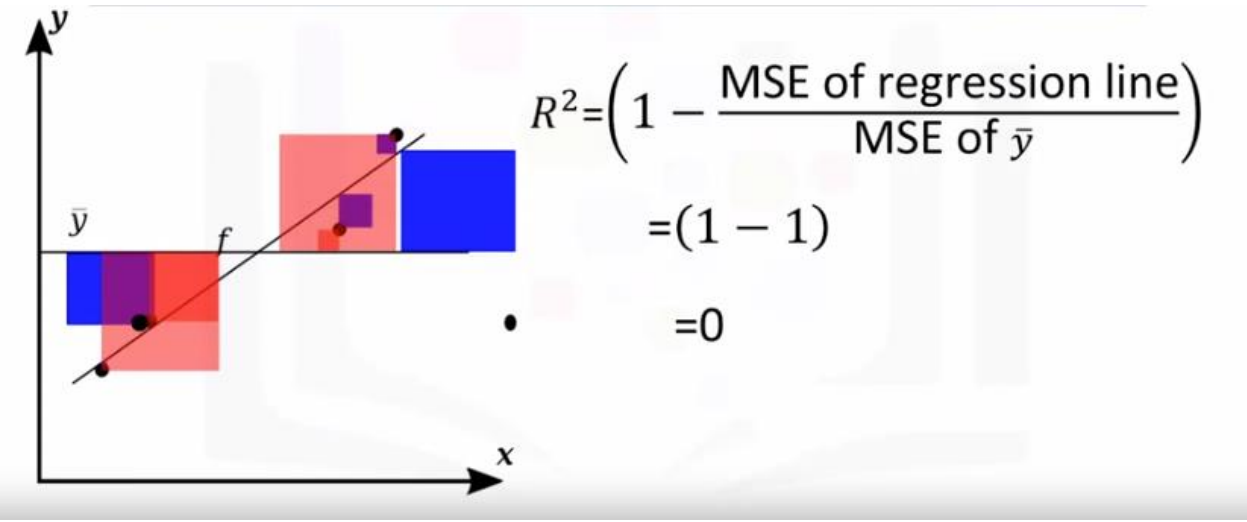
$$R^2 = \left(1 - \frac{\text{MSE of regression line}}{\text{MSE of } \bar{y}}\right)$$

$$= (1 - 0)$$

$$= (1 - 0)$$

$$= 1$$



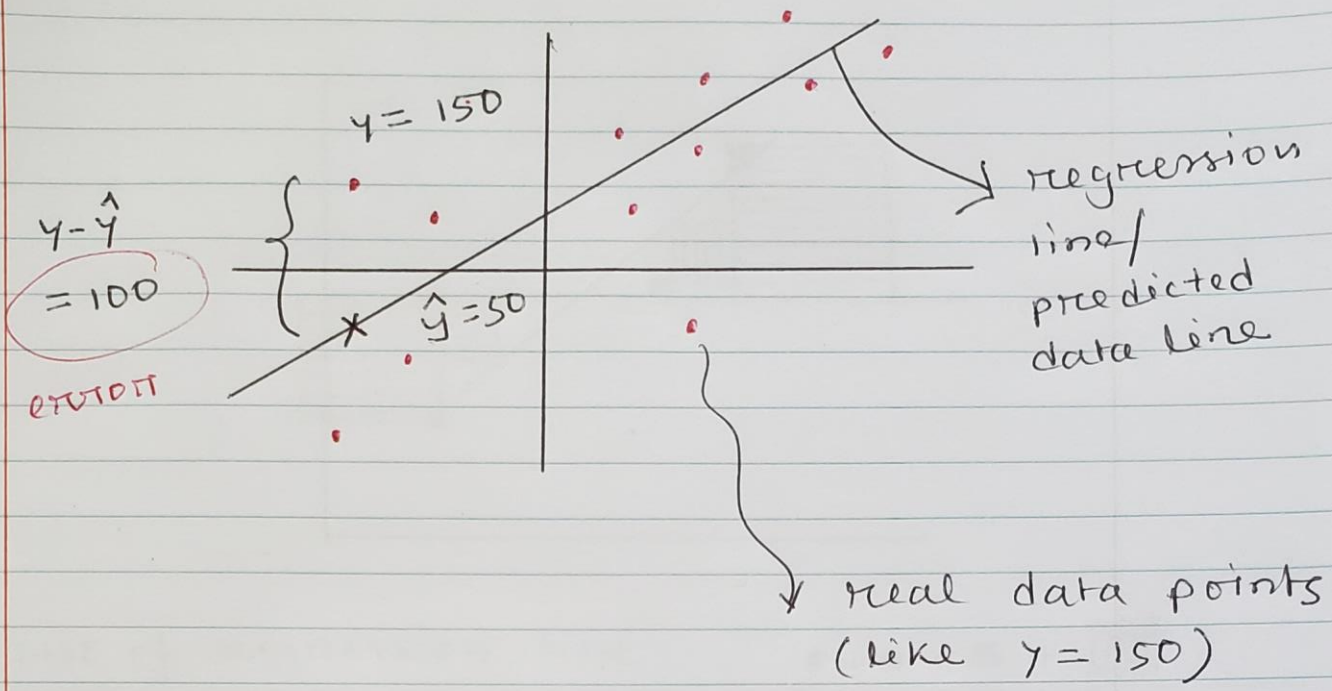


$$R^2 = \left(1 - \frac{\text{MSE of regression line}}{\text{MSE of } \bar{y}}\right)$$

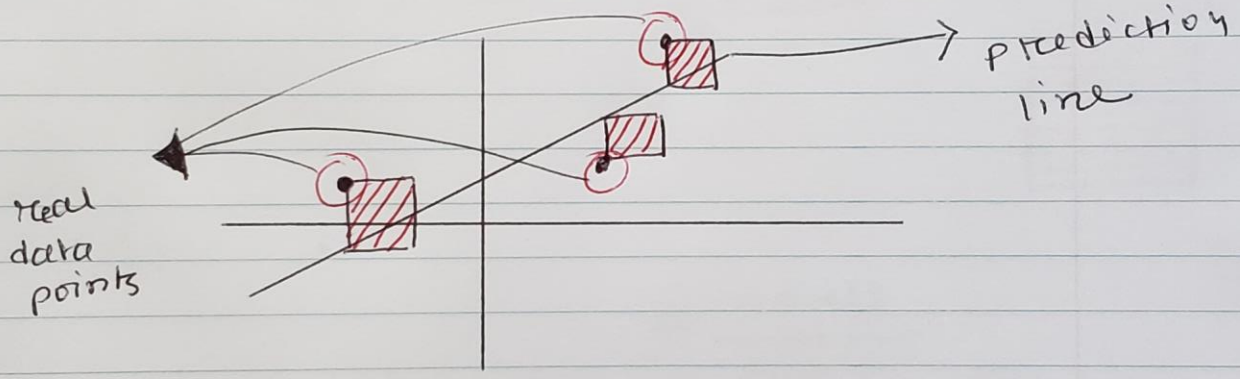
$$= (1 - 1)$$

$$= 0$$

MSE :

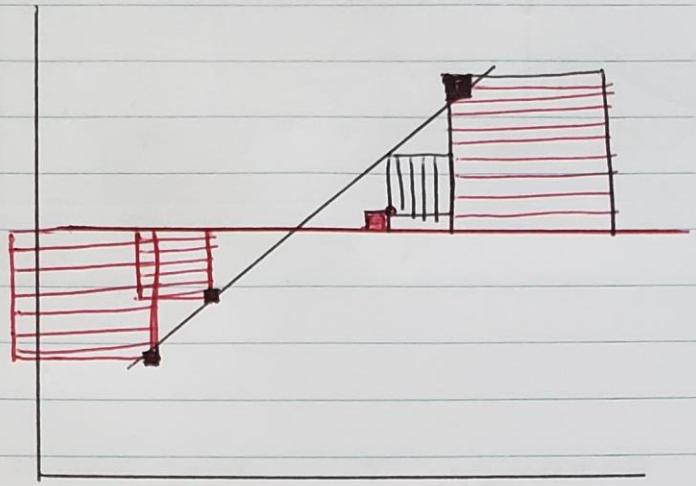


Square all the errors :



$$MSE = \frac{\Sigma \text{ all squared errors}}{\text{no. of samples}} = \frac{\square + \square + \square}{\text{no. of samples}}$$

Case 1 : Good model



$$\frac{\text{MSE of regression line}}{\text{MSE of avg of data}} = \frac{\blacksquare + \blacksquare + \blacksquare + \blacksquare}{\blacksquare + \blacksquare + \blacksquare + \blacksquare}$$

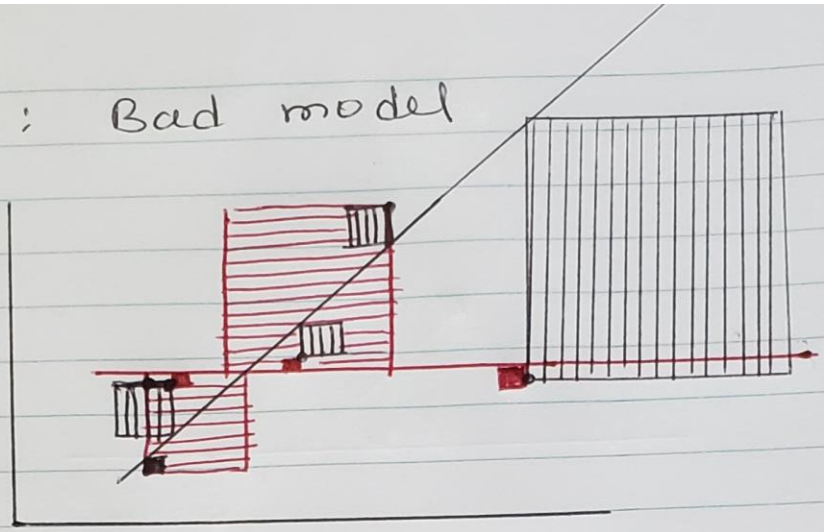
$$= \frac{\text{small}}{\text{big}}$$

$$\simeq \text{very small}$$

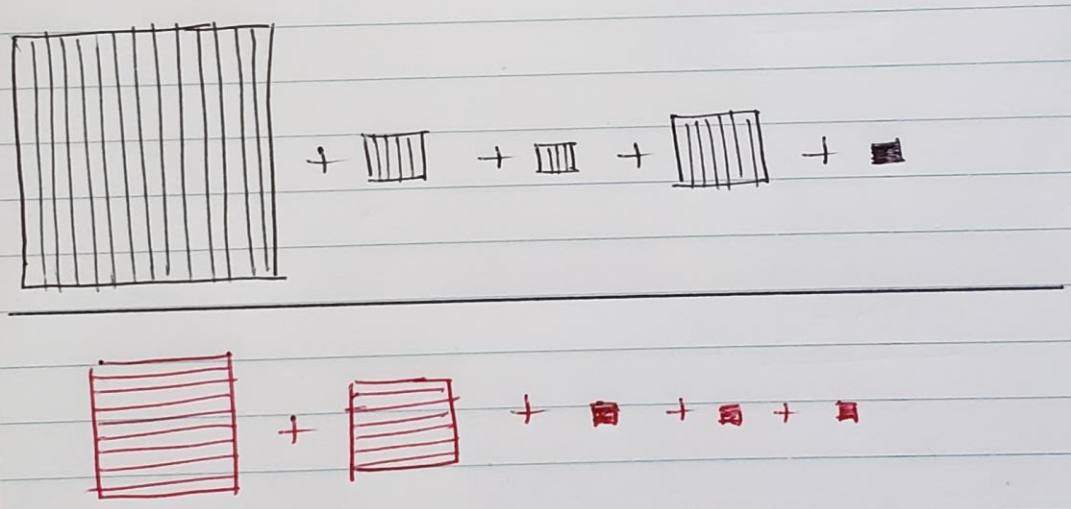
$$R^2 = 1 - \text{very small}$$

$$\simeq 1$$

Case 2: Bad model



$$\frac{\text{MSE of regression line}}{\text{MSE of } \bar{y}} = \frac{\sum \text{squared error (reg)}}{\sum \text{squared error } (\bar{y})}$$



$$= \text{Big} / \text{small} = \text{Big}$$

$$R^2 = 1 - \text{Big} \simeq 1 - 1 \simeq 0$$