_VOIS


# Regression Methods : Linear Regression & Evaluation

# Contents

- Concept of Regression
- How to identify problem is a regression problem ?
- Types of Regression
- Linear Regression
- Evaluation methods
- Hands-on

# Concept of Regression

Statistical process of estimating relation between independent and dependant variable[1]



# How to identify problem is a regression problem ?

Dependant variable (outcome) should be continuous value (numeric) not a categorical value.

e.g.

Categorical data – months of year

Continuous data – Temperature measured for a city

# Types of Regression[2]

- Linear Regression
- Logistic Regression
- Ridge Regression
- Lasso Regression
- Polynomial Regression
- Bayesian Linear Regression

# Linear Regression

- The predictor/features/ independent variables are in linear relation with dependant variable (outcome).
- If only one independent variable is present then this is called as simple linear regression.

$$Y = \alpha\_0 + \alpha\_1 . X1 + E$$

Where y = Outcome / Dependant variable

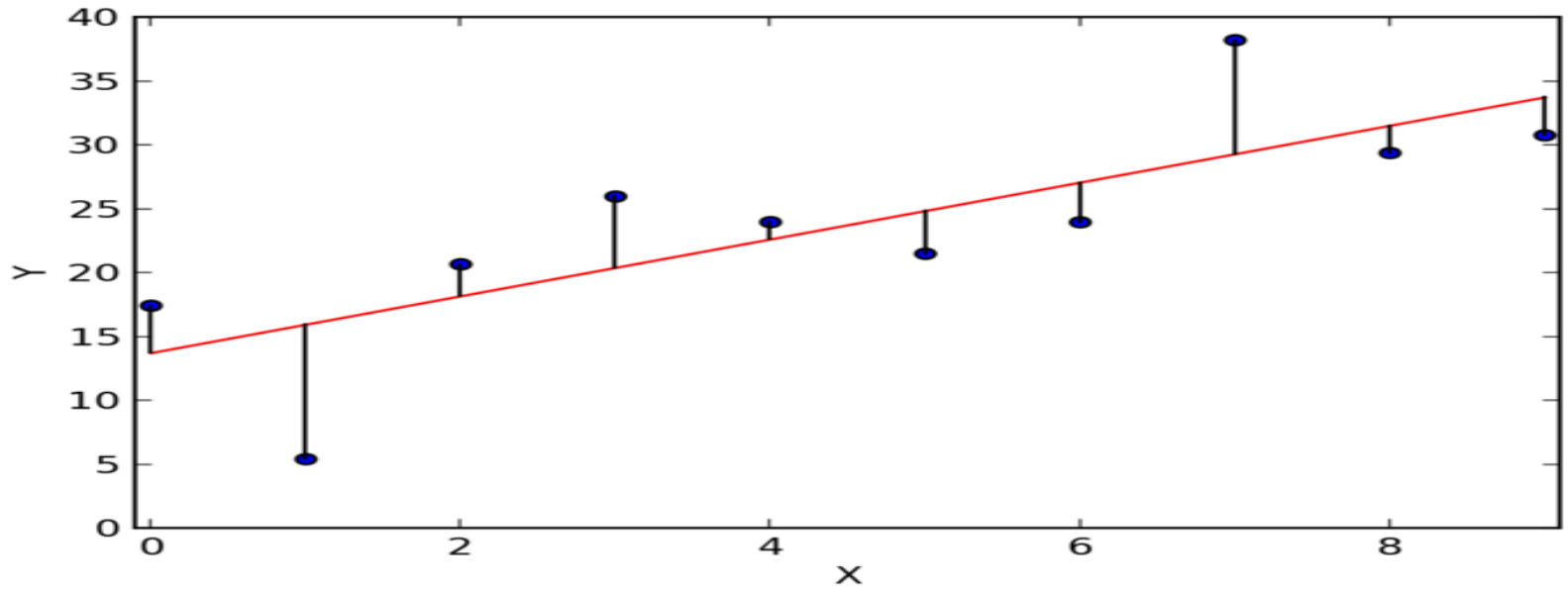
α_0 & α_1 = regression coefficients

X1 = Independent variable / predictor / feature

E = Error between predicted and original value

- If the number of independent variables are more than one then its Multiple linear regression.

$$Y = \alpha\ 0 + \alpha\ 1\ .\ X1 + \ldots\ldots + \alpha\ n\ .\ Xn + E$$




Source[3]

# Evaluation methods

- **Mean Absolute Error**

summation of the absolute value distance from the points to the line, we get Mean absolute error.[4][5][6]

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

- **Mean Squared Error**

  a summation of the square of distances from the points to the line, we get Mean squared error. [4][5][6]

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

- **R2 score**

R Square measures how much of variability in dependent variable can be explained by the model. It is square of Correlation Coefficient(R) and that is why it is called R Square[4][5][6]

$$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

(sum of squared of prediction error divided by the total sum of square which replace the calculated prediction with mean. R Square value is between 0 to 1 and bigger value indicates a better fit between prediction and actual value.)

# Hands - on

# References:


1. https://en.wikipedia.org/wiki/Regression_analysis
2. https://www.upgrad.com/blog/types-of-regression-models-in-machine-learning/
3. https://commons.wikimedia.org/wiki/File:Residuals_for_Linear_Regression_Fit.png
4. https://medium.com/acing-ai/how-to-evaluate-regression-models-d183b4f5853d
5. https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/regress.htm#CHDBHBDI
6. https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b