**Walchand College of Engineering, Sangli**

## **Machine Learning Lab (6CS372)**

**TY BTech | AY 2023-2024 | Even Sem**

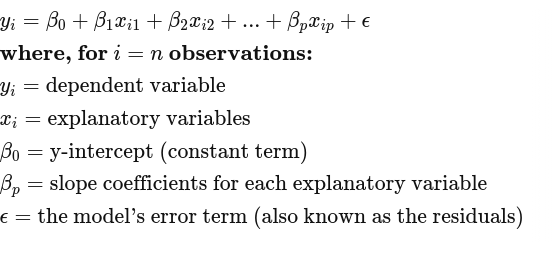
**Assignment 4**

**Name : Viraj patil**

**PRN: 21510097**

**Batch : T5**

1. **Linear regression: Single feature vs multiple features**
2. Download dataset as per your batch.
3. Preprocessing: Null value handling, standardization, replace categorical values with numeric values (e.g. 0, 1, 2 etc.)
4. Data splitting: Split data as 70% train and 30% test using train\_test\_split function.
5. Fit model using fit function taking a single feature at a time and all independent features at a time.
6. Report parameter values, training error and test error and model accuracy for Linear regression with single feature and multiple features.
7. **Answer following questions (include question and answer as markdown cell in your notebook)**
   1. Provide a general multiple linear regression equation and explain all the terms.



* 1. Explain the concept of a dummy variable and how such variables are calculated. Why is it necessary to convert nominal variables to dummy variables when performing linear regression?

A dummy variable is a binary variable used to represent categorical data in statistical analysis, where 1 indicates the presence of a category and 0 its absence. It is necessary to convert nominal variables to dummy variables in linear regression to incorporate categorical information into numerical models.

* 1. Explore and mention assumptions in linear regression with suitable explanation.

In linear regression, assumptions include linearity (relationship between independent and dependent variables), independence of errors (residuals), homoscedasticity (constant variance of residuals), absence of multicollinearity (independent variables are not highly correlated), and normality of residuals (residuals follow a normal distribution), all of which are critical for accurate parameter estimation and valid inference. Violations of these assumptions can lead to biased estimates and unreliable predictions in regression analysis.