25 Regularization Technique

L1 (Lasso Regularization) L2 (Ridge Regularization)

- Used in linear regression mostly
- This is a form of regression, that constraints/regularizes or shrinks the coefficients estimates towards zero
- This technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.
- Regularization can achive this motive with 2 techniques:

1. Ridge Regularization/L2
2. Lasso Regularization/L1

- it helps in feature selection
- it helps reducing overfitting
- it removes the data with smaller coefficients or unwanted columns or columns/data which will have neglibile impact on the final outcome

25.1 Regularization Technique (Lasoo Regularization/L1)

- This is a regularization technique used in feature selection using a **shrinkage method** also referred as the **penalized regression method**.
- Lasso regression magnitude of coefficient can be exactly zero

The cost function is defined as:

$$\text{Cost Function} = \text{Loss} + \lambda \sum_{i=1}^{n} ||w_i||$$

Loss= sum of squared residual, **lambda** = penalty, **w** = slope of the curve

- It helps in feature selection
- It makes the column (feature) zero which do not have function in the model

 No description has been provided for this image

- the black line (lambda mod(w)) shifts towards zero iteratively

25.2 Regularization Technique (Ridge Regularization/L2)

- it is called overfitting regularization technique and it redueces overfitting

Ridge regression, also known as L2 regularization, is an extension to linear regression that introduces a regularization term to redue model complexity and **help prevent overfitting**. Ridge Regression is working value/magnitude of coefficients is almost equal to zero

Its cost function is defined as:

$$\text{Cost Function} = \text{Loss} + \lambda \sum_{i=1}^{n} \|w_i\|^2$$

Loss= sum of squared residual, **lambda** = penalty, **w** = slope of the curve

 No description has been provided for this image

- it will not make exactly zero, but bring it towards zero
- L2 also reduces computational power, means reduces complexity of problem, it speeds up the model building

In [ ]: