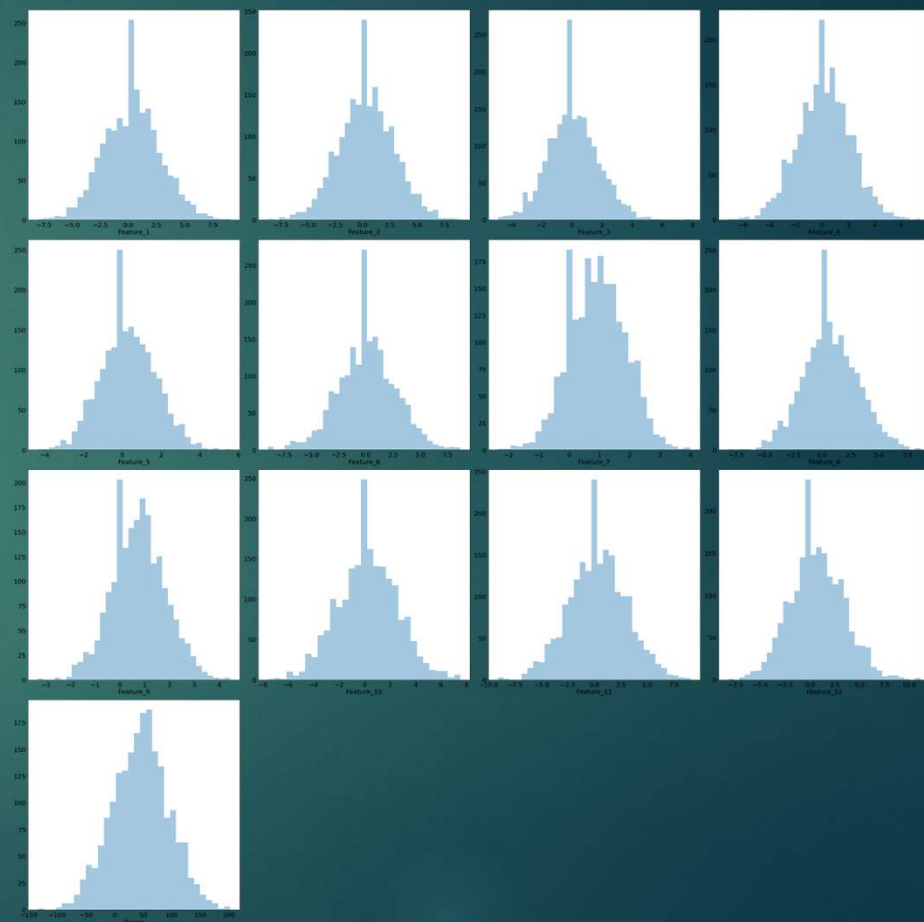
Data Modelling Steps

- Loading the data
- Missing Value Percentages Evaluation
- Imputing the Missing Values
- Analyzing numerical and categorical features
- One hot Encoding Nominal Categorical Variables
- Label Encoding Ordinal Categorical Variables
- Scripting Linear Regression (Gradient Descent, Mini Batch Gradient Descent, Stochastic Gradient Descent)

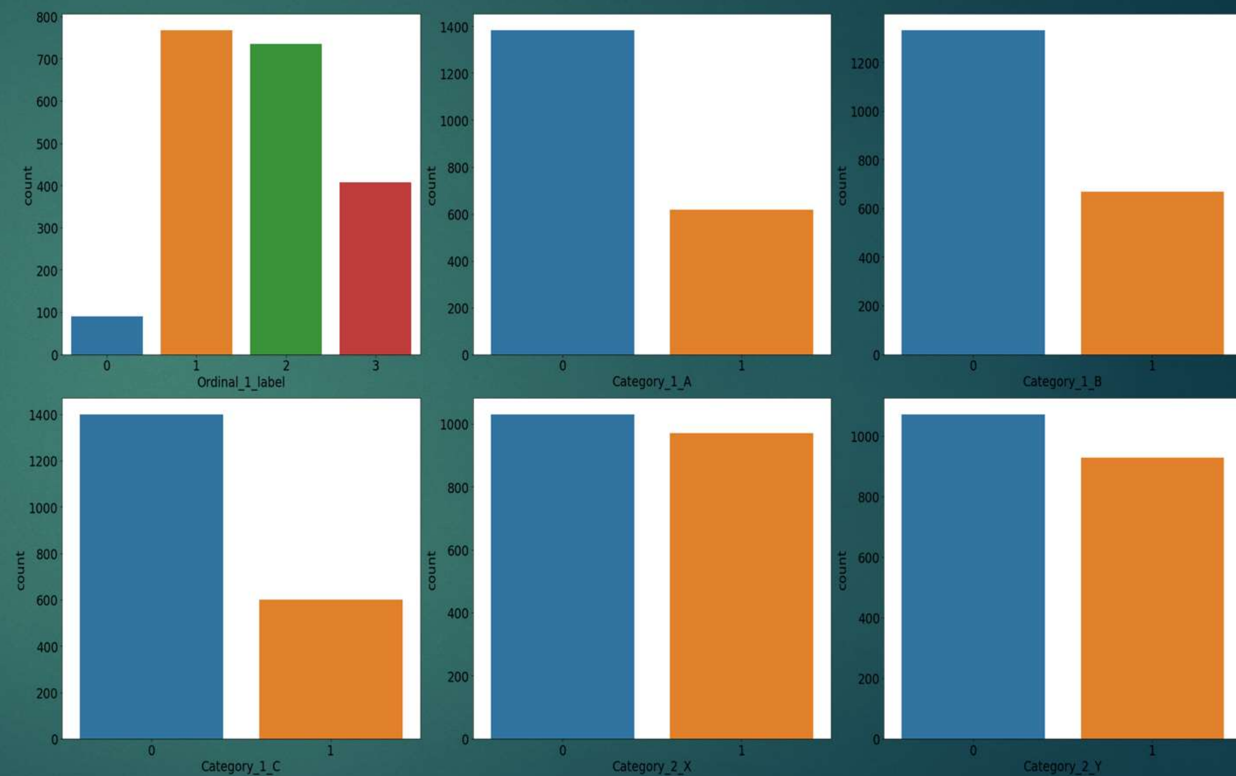
Numerical Variables Visualization

- All numerical variables follow normal distribution.
- Linear Regression assumption of normality is followed
- No zero variance features



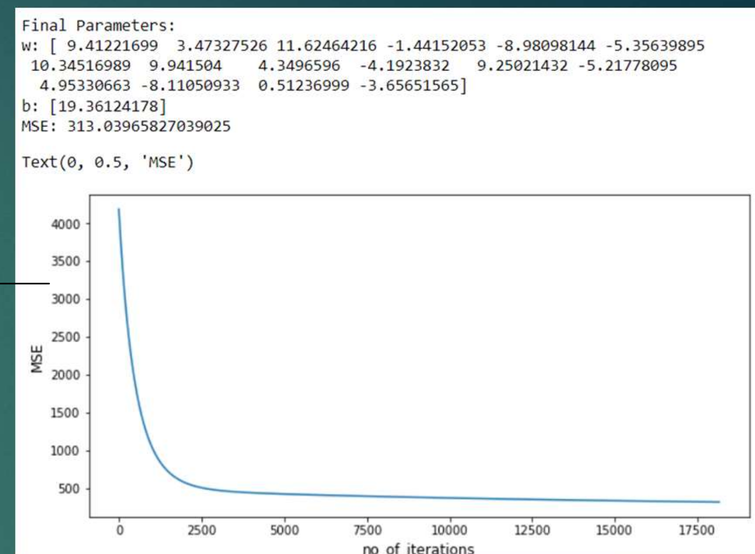
Categorical Variables Visualization

- Count Plot of ordinal and nominal categorical variables is shown after one hot and label encoding.
- No Zero Variance Features

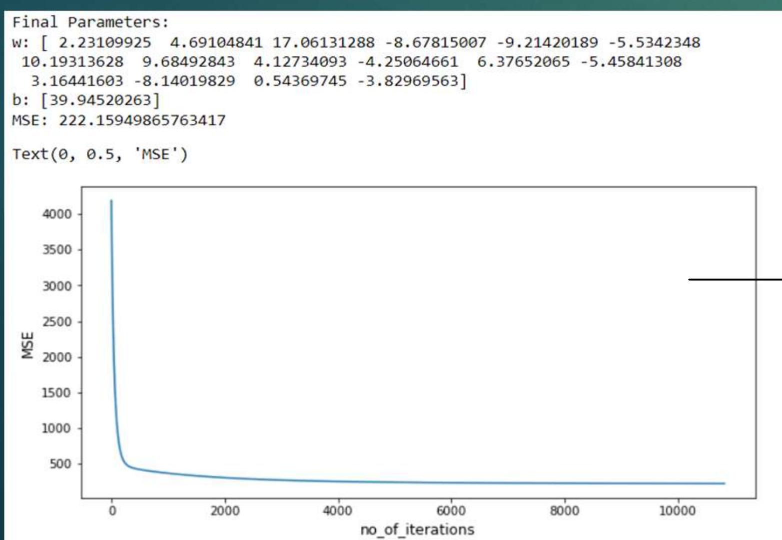


Gradient Descent Algorithm

- The plot of Mean Squared Error (J) vs iteration is shown.
- Learning Rate alpha is chosen as 0.0001 and 0.001
  Tolerance is chosen as 0.001
- If difference in the parameters w and b of successive iterations is less than or equal to the certain chosen tolerance then it means the algorithm has been converged.



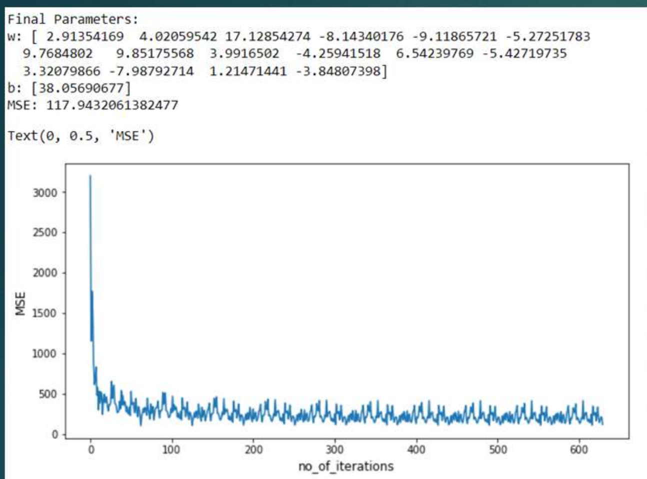
Alpha=0.0001



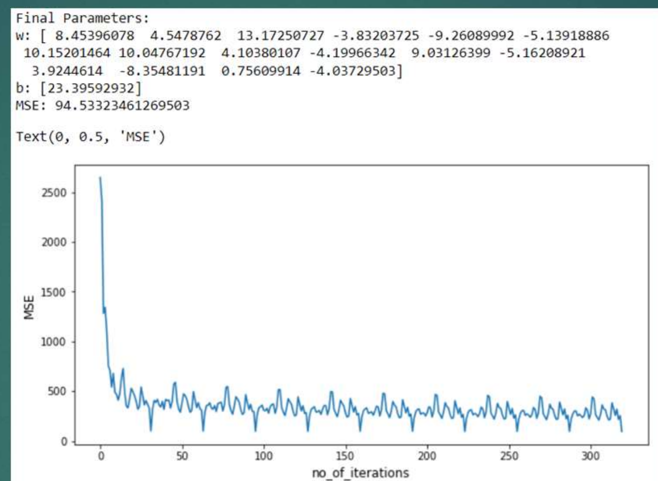
Alpha=0.001

We can see that the algorithm when run with high learning rate show a steep decrease in MSE at first in contrast to slow and gradual decrease of MSE when low value of learning rate is chosen.

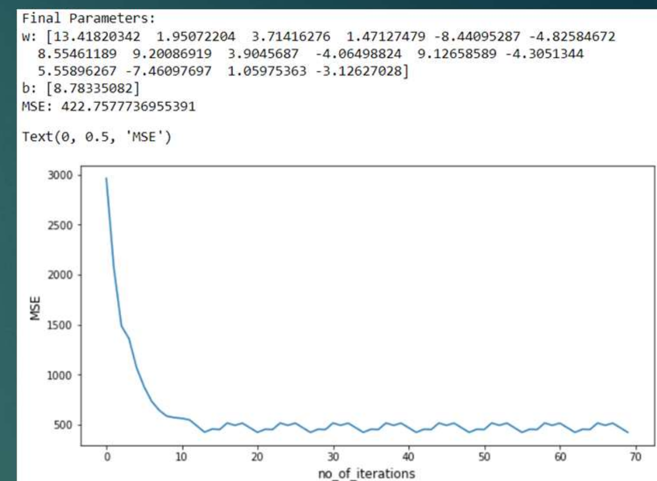
Mini Batch Gradient Descent Algorithm



Batch Size=32



Batch Size=64



Batch Size=300

Parameters Chosen :

Alpha = 0.00001

Tolerance = 0.001

Epochs = 10

Number of Observations = 2000

Observation : Gradient Descent with smaller batch sizes are noisy. Also the smaller loss is associated with smaller batch size.

Stochastic Gradient Descent Algorithm

Parameters Chosen :

Alpha = 0.00001

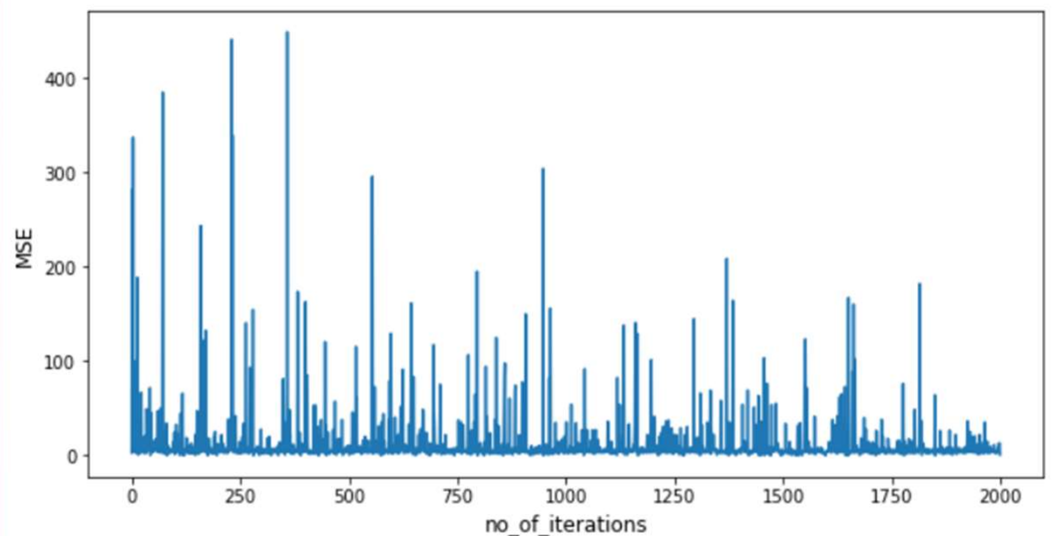
Tolerance = 0.001

Epochs = 1

Number of Observations = 2000

```
Final Parameters:
w: [ 1.49479028e+00  1.42199840e+00  1.67387797e+01 -6.83488654e+00
 -7.00805790e+00 -5.75129863e+00  9.09599829e+00  1.01360720e+01
  7.41254119e+00 -3.90446407e+00  5.35582514e+00 -5.77625766e+00
  4.60725098e+00 -9.15276786e+00 -1.52769301e-02 -1.61331164e+00]
b: [40.32257667]
MSE: 0.8373255466800653

Text(0, 0.5, 'MSE')
```



In Stochastic Gradient Descent parameter update takes at each training example. Its just a Mini Batch Gradient Descent where batch size equals to 1.

Observation :Stochastic Gradient Descent takes a very noisy path in its attempt to reach global minimum.
In fact it may never reach global minimum but it will keep oscillating around the global minimum.

L2 Regularization

With L2 Regularization we are just trying to reduce the model complexity by penalizing the parameters.

$$J = \sum_{i=1}^{m} (ypred - y)^2 \cdot \frac{1}{m}$$

With Regularization

$$J = \frac{1}{m} \cdot \left(\sum_{i=1}^{m} (ypred - y)^2 + lambda \cdot \sum_{j=1}^{p} ||\theta_j||^2 \right)$$

$$\frac{dJ}{d\theta_j} = \frac{1}{m} \left(\sum_{i=1}^{m} 2 \cdot (ypred - y) \cdot x_j^i + 2 \cdot lambda \cdot \theta_j \right)$$

Parameters Update

$$\theta j := \theta_j - \alpha \cdot \frac{dJ}{d\theta_j}$$

Putting the values we get

$$\theta_j := \left(\theta_j - 2 \cdot \alpha \cdot \frac{lambda}{m} \right) - \frac{2}{m} \cdot \alpha \cdot \left(\sum_{i=1}^{m} (ypred - y) \cdot x_j^i \right)$$

$$ypred = \theta_0 + \theta_1 \cdot x_1 + \theta_2 \cdot x_2 + \ldots\ldots\ldots\ldots\ldots \theta_p \cdot x_p$$

$$where\ m = Number\ of\ Observations,\ p = Number\ of\ Features$$

$$lambda = Regularisation\ Parameter\ and\ \alpha = Learning\ Rate$$