→) Multiple polynomial regression

Dataset :

.) m = no of training samples
.) n = no. of features

.) Choosing a different power of a feature according to their graph with target variable

1.) x1 transaction date : $\sin(x1)$
2.) x2 house age : $x2^2$
3.) x3 distance to nearest MRT station : $1/x3$
4.) x4 number of convenience stores : $x4$
5.) x5 latitude : x_5^3
6.) x6 longitude : x_6^3

→) Equations :

.) i refers to one training sample meaning one row

.) $x_i = [\sin(x1), x_2^2, \frac{1}{x_3}, x_4, x_5^3, x_6^3]$ $(1 \times n)$

.) $W = [0, 0, 0, 0, 0, 0]$ $(1 \times n)$

.) $B = 0$ (1×1)

.) $\hat{y}_i = W @ x_i.T + B$

- $L \mathrel{+}= \frac{1}{2m}(\hat{y}_i - y_i)^2$

- $\underset{(1\times n)}{\frac{\partial L}{\partial W}} \mathrel{+}= \frac{\partial L}{\partial \hat{y}_i} \times \frac{\partial \hat{y}_i}{\partial W} = \frac{1}{m}\underset{(1\times 1)}{(\hat{y}_i - y_i)}\, \underset{(1\times n)}{x_i}$

- $\frac{\partial L}{\partial B} \mathrel{+}= \frac{\partial L}{\partial \hat{y}_i} \times \frac{\partial \hat{y}_i}{\partial B} = \frac{1}{m}(\hat{y}_i - y_i)$

- $W = W - \alpha \frac{\partial L}{\partial W}$

- $B = B - \alpha \frac{\partial L}{\partial B}$

→ Findings:

- Z-Score normalization > Min Max Scaling
- Reciprocal feature was increasing the loss
- Linear regression converges faster
- Sin function does not really make a difference
- Quadratic scaling increases the error
- Making X_5 raise to power 4 reduced loss

.) After all these adjustments linear and polynomial become comparable give same MAE but polynomial has much higher loss than linear so I think it is a more generalized model because still it is giving same MAE