

Assignment 01

Learning from Data, Related Challenges, Linear Models for Regression

submitted for

EN3150 - Pattern Recognition

Department of Electronic and Telecommunication Engineering
University of Moratuwa

Udugamasooriya P. H. J.
220658U

12 August 2025

1 Impact of Outliers on Linear Regression

Question 02 We start by representing the independent in a matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},$$

and the dependent variables in a vector

$$\mathbf{y} = (y_1 \quad \cdots \quad y_n)^\top,$$

and look for a vector of weights $\mathbf{w}_{\text{OLS}} = (w_0 \quad w_1)^\top$ such that

$$\mathbf{w}_{\text{OLS}} = \arg \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})^2.$$

We directly use the result that

$$\mathbf{w}_{\text{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

We implement exactly what is described above in code, and obtain the following result:

Ordinary Least Squares Weights (\mathbf{w}): [3.91672727 -3.55727273]

Hence

$$\mathbf{w}_{\text{OLS}} = \begin{pmatrix} 3.91672727 \\ -3.55727273 \end{pmatrix},$$

and the predicted linear model is

$$y = 3.91672727 - 3.55727273x.$$

A plot of the given data points against the predicted values is shown in Figure 1.

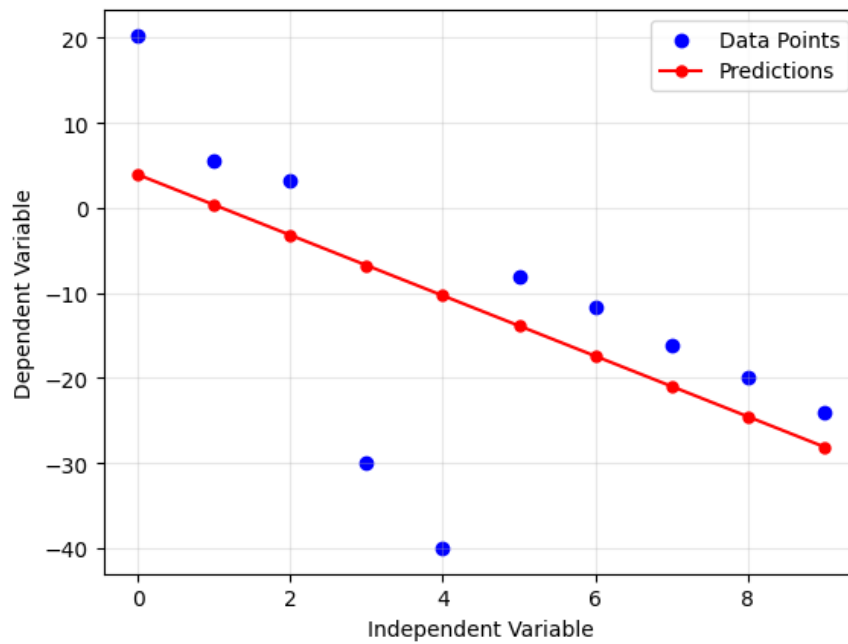


Figure 1: abc

Question 04 abc
 Question 05 pqr
 Question 06 xyz
 Question 07 def
 Question 08 ghi

2 Loss Functions

Question 01 We calculate the squared error

$$SE(\hat{y}_i, y_i) = (\hat{y}_i - y_i)^2$$

and binary cross entropy

$$BCE(\hat{y}_i, y_i) = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)$$

of each predicted value \hat{y}_i against the given corresponding target value y_i .

True Value (y_i)	Predicted Value (\hat{y}_i)	SE(\hat{y}_i, y_i)	BCE(\hat{y}_i, y_i)
1	0.005	0.9900	5.2983
1	0.010	0.9801	4.6052
1	0.050	0.9025	2.9957
1	0.100	0.8100	2.3026
1	0.200	0.6400	1.6094
1	0.300	0.4900	1.2040
1	0.400	0.3600	0.9163
1	0.500	0.2500	0.6931
1	0.600	0.1600	0.5108
1	0.700	0.0900	0.3567
1	0.800	0.0400	0.2231
1	0.900	0.0100	0.1054
1	1.000	0.0000	0.0000
	Mean	0.4402	1.4407

A plot of the different losses against the predicted values is shown in Figure 2.

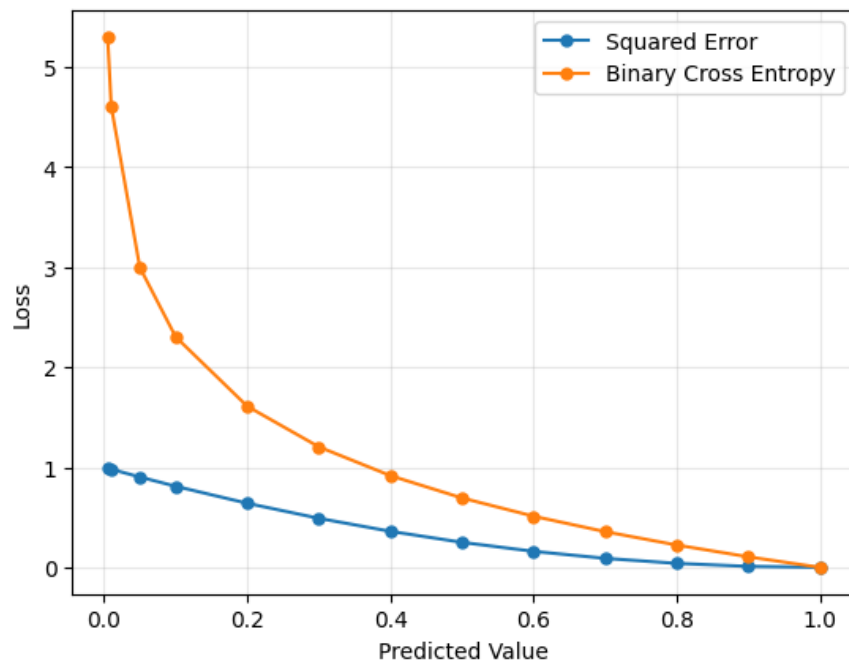


Figure 2: abc

Question 02

3 Data Pre-Processing

Question 01