Ahmet Sabri MERT
64170031

Machine Learning ProblemSet1

1.a) All necesseray libraries were imported. Then diabeties datasets were loaded. 80% of the datasets were splitted as training data and 20% of the data were splitted as a testing data. In this dataset there is 442 data, so the 80% of the 442 is calculated as 354. After splitting the datas, direct solution formula was applied and w value was calculated. This value was obtained as 304.18307453.

b) For performing gradient descent algorithm, for loop was created. In this loop, y = ax + b linear equation was written. Then, empirical risk was calculated. To minimaze the empirical risk, derivatives were taken and the new values of a and b were calculated. And all a and b values for each iterations and the empirical risk values were printed as 5734.1055150919055.

c) Here, only poltting parts were performed. In figure 1, a (w) values were plotted.In figure 2, b values were plotted .In figure 3, empirical risk values were plotted.

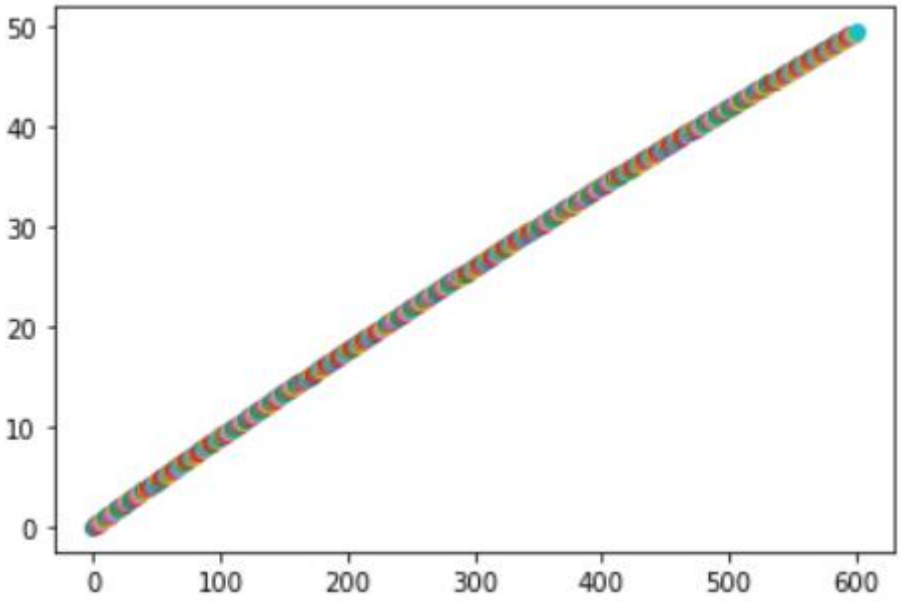

Figure 1: Plot for estimated parameter a (w) values.

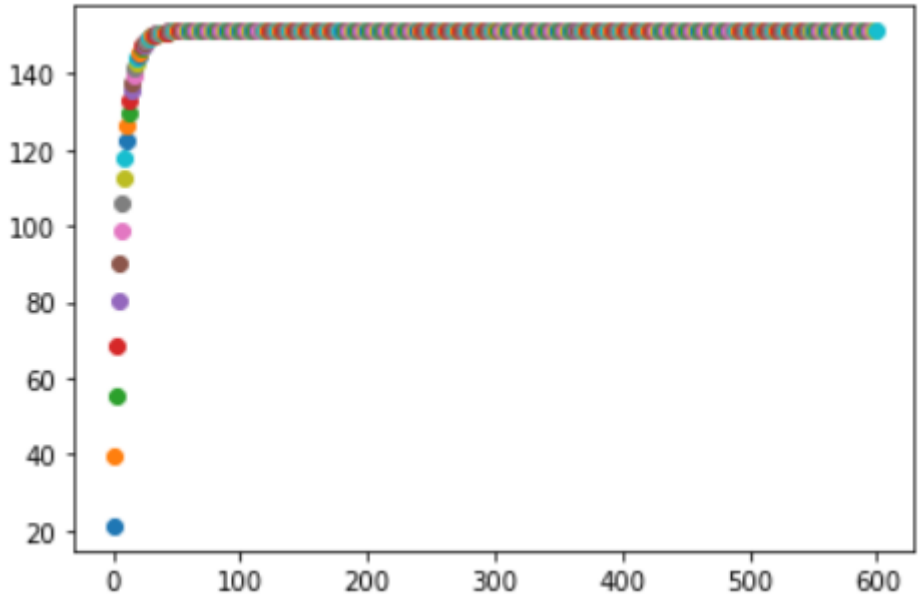

Figure 2: Plot for estimated parameter b values.

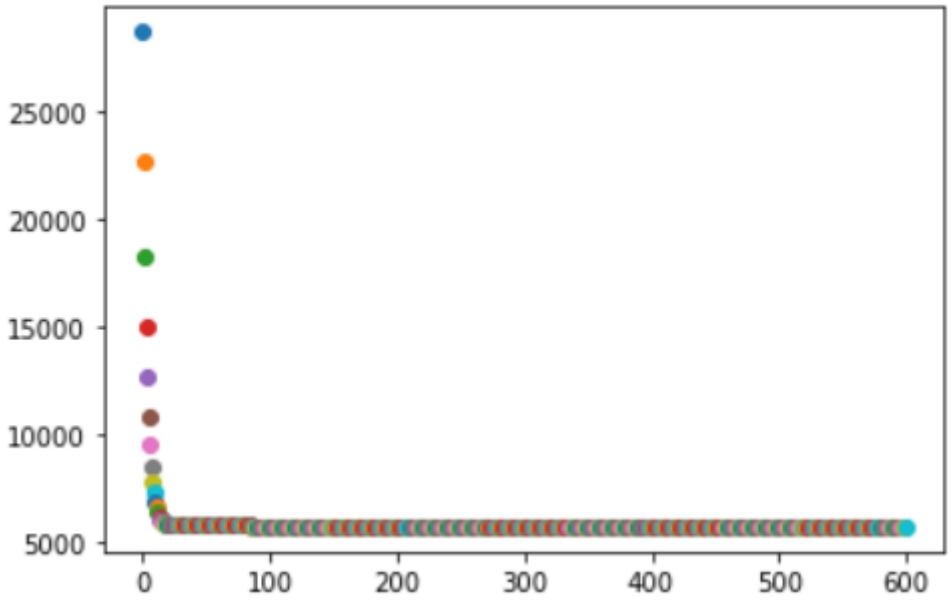

Figure 3: Plot for empirical risk values

.

d)

Here linear regression model was created. Then, coefficients, mean square error and coefficient of determination (R2) values were calculated and printed as Coefficients: 276.05986226, Mean squared error: 6164.79, Coefficient of determination: 0.05.

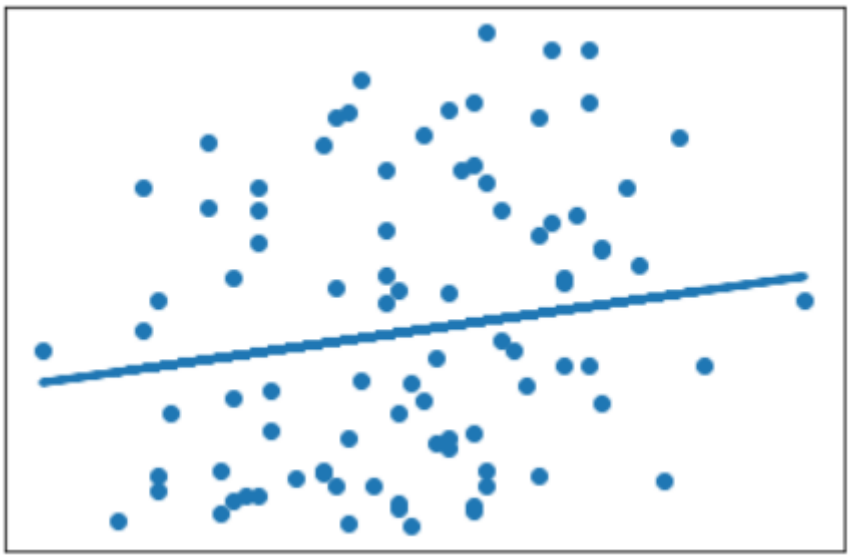

Figure 4: Plot for data points and tested regression line.

e) Here, new h(x) is written using test datasets. With these test datasets, prediction was performed and the obtained regression line was plotted. At the end, R2 score was obtained as 0.01.

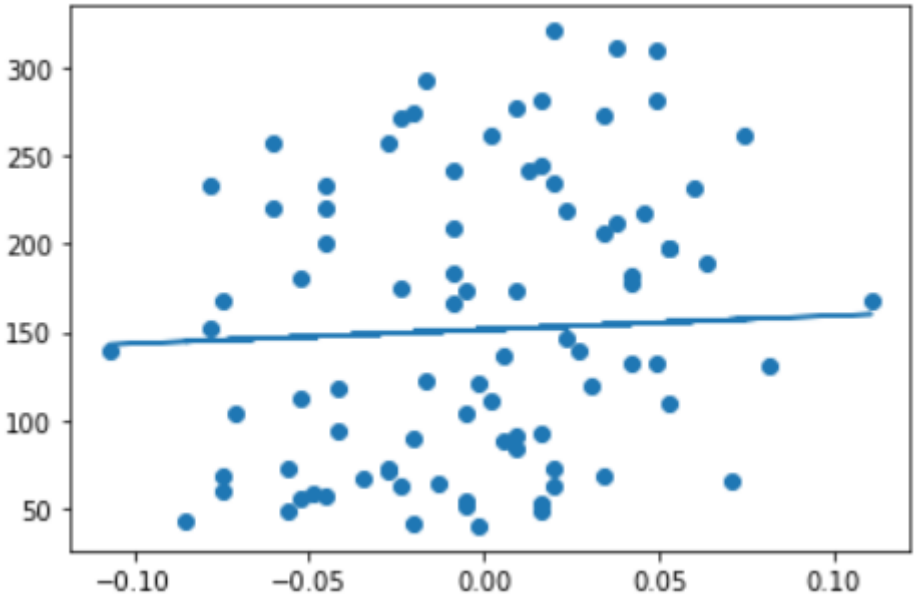

Figure 5: Plot for data points and predicted regression line.