COMP 462 Introduction to Machine Learning
Assignment 5 - Regression with Perceptron

Elif Erden 041503038

03.04.19

1 Introduction

In this assignment, a perceptron was implemented for the regression. The stochastic gradient descent algorithm was used to learn perceptron weights. The idea behind the stochastic gradient descent algorithm is to approximate the gradient descent search by updating weights incrementally, following the calculation of the error for each individual example.

1.1 Datasets

In this implementation, two different datasets were used to calculate regression equations with the perceptron. These two datasets have one feature as input variable and can be seen in Figure 1 and Figure 2.

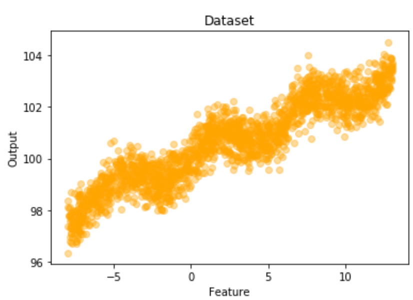

Figure 1: Dataset 1

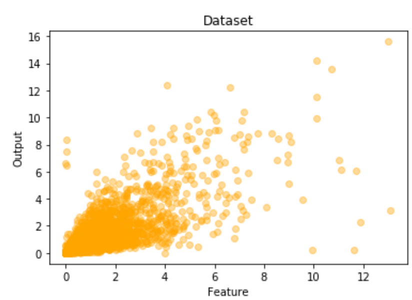

Figure 2: Dataset 2

2 Algorithm Explanation

For implementing the perceptron, firstly one column is added to the original dataset to record the output values that algorithm finds and is set to zero by default at the beginning. Then, the dataset is splitted into equal sized training and test sets. Initially, the perceptron weights are randomly selected between -0.01 and 0.01. When the stopping condition does not hold, parameters are learned using the training set according to the update equation for the perceptron: $w_i \leftarrow w_i + \Delta w_i$ where the Δw_i is calculated as $\Delta w_i = \eta(t - o)x_i$. The learning rate is set to 0.005 for the dataset 1 and 0.00005 for the dataset 2. Both sets are used to compute loss at each iteration. Loss is calculated as shown in Equation (1). If the difference between two consecutive loss value is less than 0.0001, then stopping condition holds and algorithm stops. Finally, plot of the data with regression equation and loss function for both test and training sets are plotted.

$$E[\vec{w}] = \frac{1}{2} \sum_{d \epsilon D} (t_d - o_d)^2 \tag{1}$$

3 Results

The algorithm was tested on two different datasets. The scatter plot of the data with the fitted regression line for the dataset 1 can be seen in Figure 3 and the loss function of the dataset 1 for both test and training sets can be seen in Figure 4. From the loss function, it can be seen that training and test set errors are very close to each other so that they cannot be distinguished in the plot. The fitted regression equation for the dataset 1 is shown in Equation (2).

$$o = 99.941 + 0.193x_1 \tag{2}$$

The scatter plot of the data with the fitted regression line for the dataset 2 can be seen in Figure 5 and the loss function of the dataset 2 for both test and training sets can be seen in Figure 6. From the loss function, it can be seen that training and test set errors are very different so that they can be easily distinguished in the plot. The fitted regression equation for the dataset 2 is shown in Equation (3).

$$o = 0.249 + 0.813x_1 \tag{3}$$

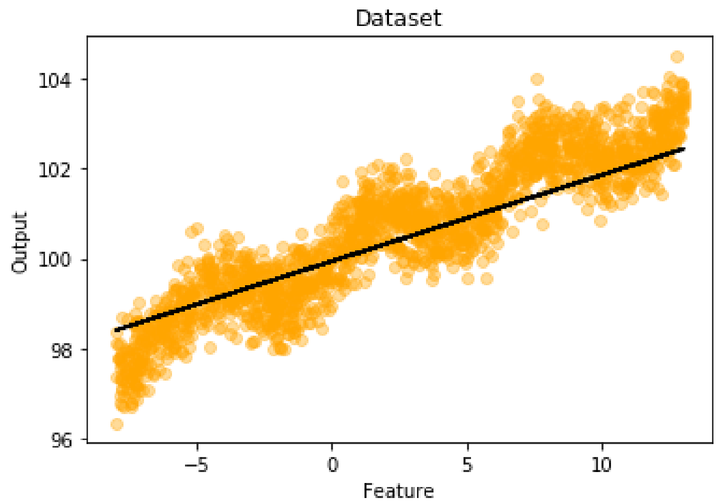

Figure 3: Dataset 1 with the regression line

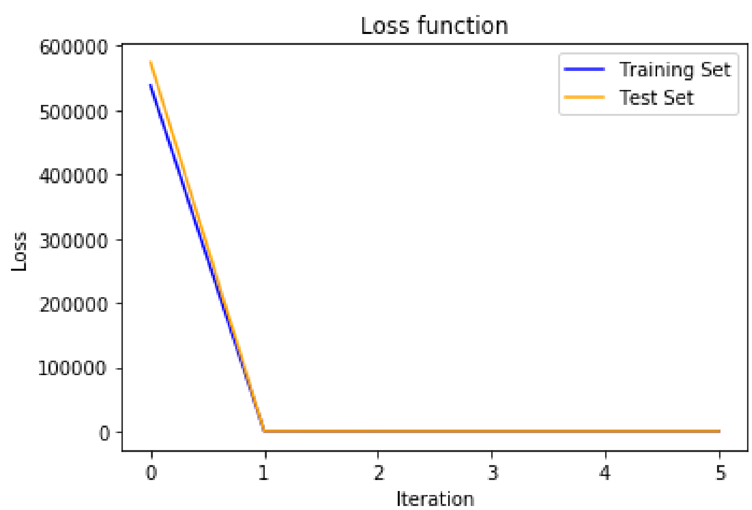

Figure 4: Loss function for dataset 1

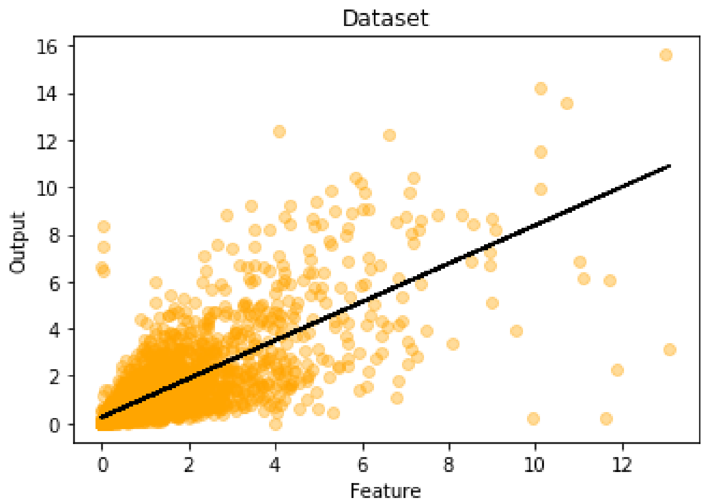

Figure 5: Dataset 2 with the regression line

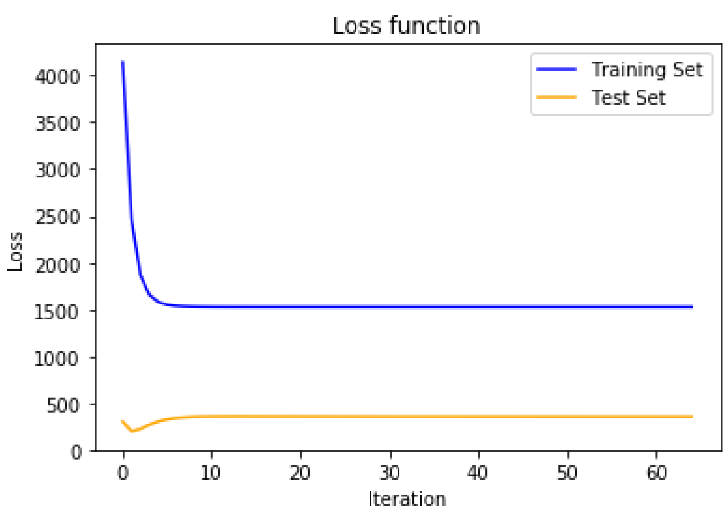

Figure 6: Loss function for dataset 2

Table 1: Pivot table for comparison of results

	Dataset 1		Dataset 2	
	w_0	w_1	w_0	w_1
Gradient Descent	99.941	0.193	0.249	0.813
Excel Scatter	99.961	0.246	0.149	0.814
Difference	0.02	0.053	0.1	0.001

After fitting regression lines for the datasets, both datasets were put into Excel to compare results. Scatter plots were plotted with displaying regression equations on the charts. The pivot table for comparing algorithm results with Excel results are shown in Table 1.

3.1 Analysis with Different Learning Rates

For analyzing regression outputs that are found by the perceptron with mentioned learning rates, different learning rates for both datasets were tried and results were analyzed. For the dataset 1, four different learning rates were tested and four regression lines for different learning rates with weight parameters can be seen in Figure 7, 8, 9 and 10. It can be seen in the plots that when the learning rate increases, there is a small decrease in the intercept and slope of the regression line.

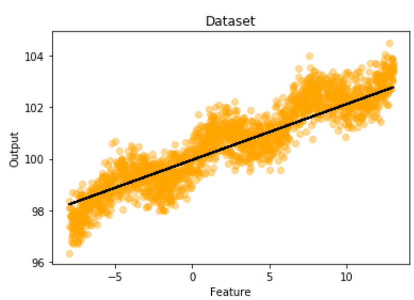

Figure 7: Dataset 1 with the regression line ($\eta = 0.003, w_0 = 99.959, w_1 = 0.217$

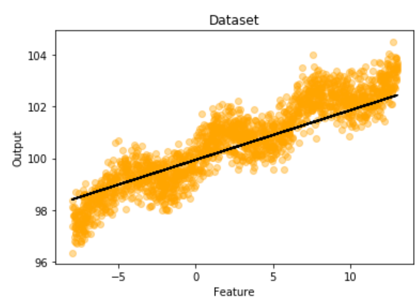

Figure 8: Dataset 1 with the regression line ($\eta = 0.005, w_0 = 99.942, w_1 = 0.193$)

For the dataset 2, four different learning rates were tested and four regression lines for different learning rates with weight parameters can be seen in Figure 11, 12, 13 and 14. It can be seen in the plots that when the learning rate decreases, there is a significant increase in the slope of the regression line.

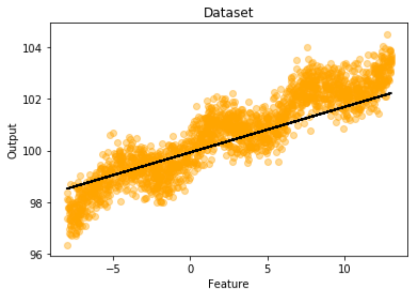

Figure 9: Dataset 1 with the regression line ($\eta = 0.007, w_0 = 99.931, w_1 = 0.177$)

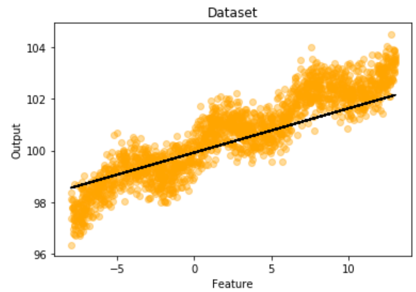

Figure 10: Dataset 1 with the regression line ($\eta = 0.009, w_0 = 99.923, w_1 = 0.172$)

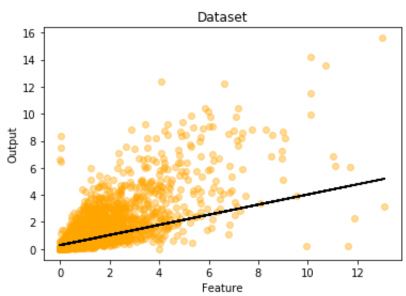

Figure 11: Dataset 2 with the regression line ($\eta = 0.05, w_0 = 0.281, w_1 = 0.375$)

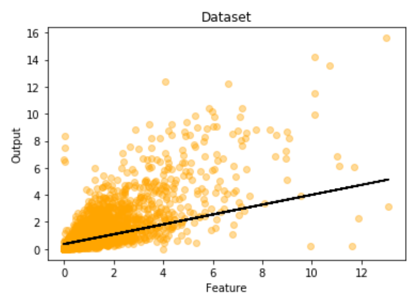

Figure 12: Dataset 2 with regression line ($\eta = 0.005, w_0 = 0.359, w_1 = 0.365$)

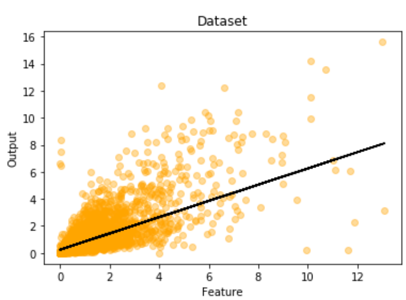

Figure 13: Dataset 2 with the regression line ($\eta = 0.0005, w_0 = 0.245, w_1 = 0.601$)

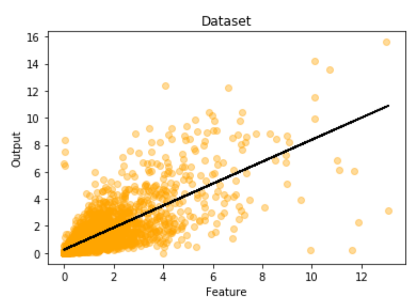

Figure 14: Dataset 2 with regression line ($\eta = 0.00005, w_0 = 0.249, w_1 = 0.813$)