CS4372 Project 1 Report

Salma Khalfallah SMK210009
Hana Al-Jaser NetID

September 21, 2025

1 Introduction

Linear regression is a simple approach to supervised learning that simply assumes a linear relationship between the predictor and the response variables in a model. For example, $f(x) = 2x$ indicates that the response variable, $y = f(x)$, is twice the value of the predictor variable x. Although a simplistic model with strong assumptions, this model is a powerful model with high-inference properties.

$$f(x) = w_0 + w_1 x_1 + \cdots + w_p x_p + \epsilon$$

An issue arises in this methodology in that is simply difficult to find the true function f of any relationship between response and predictor variables. This raises a motivation to *estimate* what the function f might look like through its weights $[\hat{w_0} \cdots \hat{w_p}]$. In the following report, our goal is to delve into two different ways to perform this estimation (Stochastic Gradient Descent, Ordinary Least Squares) and compare the results that stem from the two methodologies.

2 Ordinary Least Squares

2.1 Introducing Ordinary Least Squares

In linear regression, our objective is often to minimize the error mean squared error such that:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2$$

where the MSE simply represents the difference between the estimated response value and the true response value. Intuitively, the smaller the MSE, the lower the error in the model. By minimizing the error function, we approximate the true function f.

One way in which this problem could be solved is by using simple calculus: calculating the partial derivative of the MSE with respect to weights $w_0, \cdots, w_p$

and setting that derivative equal to zero. The intuition lies in that, when the partial derivative equals zero, the MSE would be at its minimal value. Therefore, the weights could be solved for using algebra.

$$\frac{\delta MSE}{\delta w_0} = 0, \cdots, \frac{\delta MSE}{\delta w_p} = 0$$

This solution is called the "Ordinary Least Squares" solution. Although this is a relatively slower solution ($O(p^3)$, this solution will return a *global* solution. This means that we know that MSE from this solution will be the lowest in the error space.

2.2 Pre-Processing

2.3 Model Construction

2.4 Result Analysis

3 Stochastic Gradient Descent

3.1 Introducing Stochastic Gradient Descent

Recalling our linear regression learning objective, we want to minimize the mean squared error formula (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2$$

The previous solution, although a global solution, is a very slow computational solution. For larger datasets, the OLS solution is simply not viable and requires an alternative: this is where we pivot into the Stochastic Gradient Descent algorithm.

The idea for Gradient Descent is to simply take iterative steps in the opposite direction of a function f at any point in the space. By initializing the weights $(w_0, \cdots, w_p)$ to random values and then looking for the local minimum of a specific MSE curve through an iterative function:

$$w_p^{new} = w_p^{old} - \mu(\frac{\delta MSE}{\delta w_p})\forall p$$

where w_p is the weight for predictor p, μ is a pre-assigned learning rate, and $\frac{\delta MSE}{\delta w_p}$ is the gradient of the error function with respect to the weight w_p.

This algorithm does not guarantee a global minimum by any means; gradient descent is simply going down a specific curve on the MSE function landed on through the randomization of weights. However, this is an almost linear function that works to mitigate the previous time-complexity problem arising from the OLS solution.

3.2 Pre-Processing

3.3 Model Construction

3.4 Result Analysis

4 Conclusion