CS566 Project Proposal

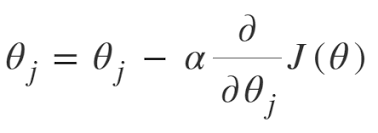
Assiya Karatay, Ayomide Awojobi, Ruiqi Chang

**Gradient Descent Algorithms**

The purpose of this project is to investigate the Gradient Descent Algorithm. This algorithm was invented by famous French mathematician Augustin-Louis Cauchy to solve quadratic problems in astronomy (Cauchy, 1847).

Gradient descent is an easy to understand and implement popular optimization strategy used in machine learning and deep learning. Gradient descent is used in machine learning to find the values of a function’s parameters which are basically the coefficients that minimize a cost function as far as possible. Gradient descent is not only limited to machine learning and deep learning. It is used widely in other areas such as control engineering(robotics, chemical, etc), computer games and mechanical engineering (Kwiatkowski, 2021).

The basic concept of the Gradient Descent Algorithm can be explained in a simple way. It is like people go down from a mountain to the lowest point , and the easiest way is to follow the gradient which is the steepest slope down the mountain. This is the gradient descent formula. When the gradient descent algorithm is applied, this formula should be repeated until convergence. The α means the learning rate which represents the size of the next step, and the right part of the α represents direction. Thus, it is important to get the appropriate α and the right direction through iterating data points in a large database (Pandey, 2019).



The steps of Gradient Descent Algorithm include choosing the initial point and finding the gradient of it. Then we make a scaled step opposite to the direction of the gradient for the minimum or in the direction of the gradient for the maximum. We repeat these steps until we get the optimum (Kwiatkowski, 2021).

In this project, we would be looking at different types of Gradient Descent Algorithms: Stochastic, Batch, and Mini-Batch Gradient Descent. The difference within these types is that Stochastic Gradient Descent takes one example in each iteration. Mini-Batch Gradient Descent takes a random sample in predetermined size while the Batch one takes the whole set in each iteration. In the last one, the batch size is equal to the data set size. With these different types of algorithms, we would compare them for design and complexity.

**Resources Used for Initial Investigation**

Cauchy, A. (1847). 1. Methode generale pour la resolution des systemes d’equations simultanees. *Comptes rendus hebdomadaires des séances de l’Académie des sciences,* 536-544.<https://gallica.bnf.fr/ark:/12148/bpt6k2982c.image.f540.pagination.langEN>

Donges, N. (n.d.). *Gradient descent: An introduction to 1 of Machine Learning's most popular algorithms*. Built In. Retrieved February 14, 2022, from<https://builtin.com/data-science/gradient-descent>

Kwiatkowski, R. (2021, May 24). *Gradient descent algorithm - a deep dive*. Medium. Retrieved February 14, 2022,from [ttps://towardsdatascience.com/gradient-descent-algorithm-a-deep-dive-cf04e8115f21](https://towardsdatascience.com/gradient-descent-algorithm-a-deep-dive-cf04e8115f21)

Pandey, P. (2019, March 18). *Understanding the mathematics behind gradient descent.* Medium. Retrieved February 21, 2022, from <https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e>