Implementing Optimization Techniques on Neural Network Backpropagation

Individual report by Aditya Nayak

I. Introduction:

Neural networks are a type of machine learning model that is inspired by the structure and function of the human brain. They consist of layers of interconnected nodes, or neurons, that process information and make predictions based on input data. Neural networks have become increasingly popular in recent years due to their ability to learn complex patterns and relationships in data. They have been applied to a wide range of fields, including image and speech recognition, natural language processing, and even game playing.

Optimization algorithms are a key component of neural network training, as they are used to adjust the weights and biases of the network to minimize the error between the predicted output and the actual output. There are several types of optimization algorithms used in neural networks, including gradient descent, stochastic gradient descent, and Adam.

II. Project work Description:

To execute our project, the tasks assigned to me were as follows:

a. Responsible for dataset collection.
b. Gather three datasets.
c. Make the datasets in proper format and ready to implement on the network.
d. Implement the variable learning rate algorithm using momentum on the neural network optimization.

1. Dataset collection

For data collection I decided to consider dataset that had at least more than 1000+ observations as they will be enough for implementing on our neural network. And as instructed by Prof. Amir Jafari I decided to select up to three datasets from Kaggle based on the number of features characterizing them as small, medium, and large.

A. Housing dataset: Small

This dataset is based on data from the 1990 California Census, and includes metrics such as the population, median income, and median housing price for each block group in California.
Block groups are the smallest geographical unit for which the US Census Bureau publishes sample data., and each block typically has between 600 to 3000 people. With the help of these features, we can predict the price of each house using our network model.

B. Spotify-YouTube dataset: Medium

Dataset of songs of various artist in the world and for each song is present. There are several statistics of the music version on Spotify, including the number of streams & the number of views of the official music video of the song on YouTube.

It includes 26 variables for each of the songs collected from Spotify. For our project, we considered 17 variables which will help us in determining the number of views/likes a song could probably get using our neural network.

C. Heart cancer dataset: Large

The following dataset consists of 34 variables along with a target variable target_deathrate at various locations throughout . No further description of the data variables was provided on Kaggle.

The dataset initially was not ready for use in the network, consisting of noisy data and unrequired features which would not contribute towards the working of the neural network. Thus, by performing data cleaning and data pre-processing I ensured that it was clean and could be efficiently used in the network.

2. Variable Learning rate:

Variable learning rate refers to an optimization technique where the rate at which a model learns from training data is adjusted dynamically during the training process. This approach is often used in conjunction with gradient-based optimization algorithms such as stochastic gradient descent (SGD) or Adam to improve their performance.

The idea behind variable learning rate is to use a higher learning rate at the beginning of training when the parameters of the model are far from their optimal values and the gradients are typically larger. As the training progresses and the parameters get closer to their optimal values, the learning rate is gradually decreased to ensure that the optimization process does not overshoot the minimum of the loss function.

Despite the successful implementation of data collection, data cleaning, and data pre-processing techniques, I encountered difficulties in implementing the variable learning rate optimization algorithm. Despite several attempts to implement the algorithm, I was not able to deliver a functioning solution due to unforeseen technical challenges. By this, I recognized the importance of this optimization technique and acknowledge the need for further research and development in this area. In future work, I intend to continue exploring optimization techniques and to build upon the knowledge gained from this project to improve the performance of machine learning models.

III. Results:

The data collection, data cleaning, & data pre-processing steps were critical to the success of our machine learning project. The careful selection of data sources and the implementation of rigorous data cleaning techniques ensured that the datasets were accurate, complete, & free from errors. The data pre-processing steps, which included feature scaling, and feature selection, were also crucial in optimizing the performance of the machine learning models.

IV. Summary:

In conclusion, for the project, I successfully implemented data collection, data cleaning, and data pre-processing techniques to prepare the datasets for use in machine learning models. These steps were essential in ensuring that the data was accurate, complete, and free from errors, which is crucial for obtaining reliable results.

Additionally, the project attempted to implement variable learning rate optimization algorithms in the neural network, but unfortunately was unsuccessful in achieving this goal for our model performance. Despite this setback, the project provided me with valuable insights into the importance of data preparation in machine learning and highlights the need for further research into optimization techniques that can improve model performance. Overall, this project demonstrates the importance of careful and thorough data preparation in the development of successful machine learning models.

Percent of code from internet = ((4 – 2)/4)*100 = 50%

V. References:

a. “Neural Network Design” (2nd Ed), by Martin T Hagan, ISBN 0971732116
b. Henry P. Gavin (2022), “The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems”, Duke University, Department of Civil & Environmental Engineering.
c. Rauf Bhat (2020), “Gradient Descent with Momentum”, Towards Data Science.
d. I. Khan et al (2020), "Design of Neural Network with Levenberg-Marquardt and Bayesian Regularization Backpropagation for Solving Pantograph Delay Differential Equations" in IEEE Access, vol. 8, pp. 137918-137933, doi: 10.1109/ACCESS.2020.3011820.
e. Bruce, P., et. al.,(2020). Practical Statistics for Data Scientists, O’Reilly