Web-based Machine Learning Application

Progress Report

GENE 191

October 30th, 2018

TABLE OF CONTENTS

1. Introduction
2. Summary
3. Discussion
   1. Work Completed
   2. Problems and Solutions
   3. Future Work
   4. Cost
4. Conclusion
5. **Introduction**

This progress report covers from the time period of September 26th to October 30th. This project is under developing by Chengyang Li. The main purpose of the build a web-based machine learning tool is to provide the public an easy access to machine learning and data optimize. A secondary purpose is to introduce people who are interested in machine learning in this area.

**2.0 Summary**

Preliminary preparation and parts of this project have been completed. The first two steps of learning and making schedule have been finished, and the back-end developing is in the process which is also where the problem occurred. One algorithm is needed to be chosen to continue this project. After that, we need to go through the final test and the public test to finish this project.

**3.0 Discussion**

**3.1 Work Completed**

Preliminary preparation and parts of this project have been completed. Firstly, I finished the study of courses of machine learning and part of the course of deep learning presented by Andrew Ng on Coursera. After that, I have a thorough understanding of contemporary machine learning and artificial neural network. I decided to use algorithms to provide the function of linear regression to the users. It is a good model for them to use and let them get attracted. The second part of this project which is making a schedule is also completed. The step of developing this project has been divided into two parts: back-end developing and front-end developing.

**3.2 Problems and Solutions**

In the part of back-end developing, there are two algorithms which can be chosen to used in this project. The first algorithm is gradient descent. The advantage of this algorithm is that it is suitable for big amount of data such as thousands of examples with thousands of features. However, the disadvantages are that it usually takes several hours for the computer to return the result, and the users need to set some hyperparameters so that it can run at the speed not too slow. The second algorithm can be used is the normal equation. It is much faster than the gradient descent, and it can always return the best result. Unfortunately, it is not suitable for a large dataset, and there are some recursions in the algorithm, which meanings the server need to make lots of computation, and it’s possible that the server suddenly runs out of memory. After comparing both of them, I decided to use the normal equation as the algorithm because the users are less likely to input a large amount of data since they are new to machine learning. Also, I can set some limitation in the part of front-end developing so that if they want to have some large calculation, they can email me, and I can help them to get their result on my computer.

**3.3 Future Work**

There is some work remains to be completed. The first remaining task is the front-end developing. It is necessary to provide the users with good user experience; however, I am not good at designing and building websites. The interface might be ugly and simple, but I believe the user can easily use it. After it, we need to have the final test and public test. The program will be improved base on the suggestions from the test. Finally, if there is no error, the program will be finished and published.

**3.4 Cost**

There is no significant cost till now. All the parts, including learning and building, are free. There is no discrepancy between the budget and reality.

**4.0 Conclusion**

This project is currently on schedule, and a few problems have been solved in order to complete this project. If there is no unusual delay, this project can be completed by the deadline.