**Final Project CS549**

**Neelam Singh Dhandaria**

**Project Topic:**

The project is done on the topic of unsupervised neural network simple competitive learning. The topic is about Student performance. Education is very important for everybody, also to know how you are performing we use the data from the real work sources and integrate them to get the best out of this so called new oil of the new world. Keeping all this in mind I have chosen to do training of simple competitive learning on the topic of student performance by getting the data from UCI which will be discussed further in this document. This project work is done for the final project for the subject CS549. The topic chosen have also been discussed in the class by our professor during the fall 2020 semester. The programs process the data of number of students and categorize them in to three different groups. Each group will have the dataset of the student who have similar type of performance as a student. This way professors or the care takers of the students can focus more on a group of students and can give more competitive challenges to another group with better performance.

**Project DataSet:**

Dataset used for the project is Student performance dataset which I have taken from UCI Machine learning repository. Link for the dataset I have used is: <https://archive.ics.uci.edu/ml/machine-learning-databases/00320/> . There are total 32 attributes from which I have taken only what I needed for this project. I have taken 8 attributes from 32 attributes and they are : extra educational support, family educational support, extra paid classes, extra curricular activities, attend nursery school, willing to take higher education, internet access at home. The type of the attributes are binary so for yes it is 1 and for no it is 0. The data is basically originated from two Portuguese schools two data was provided for performance for math and Portuguese language but I have taken only mat sub data for my project. There are total 395 datasets provided for training.

**Project Description:**

The programming language used to develop the project is java and was programmed using one of the IDE which support java programming called Netbeans. I did not implement any of the machine learning libraries in my project. First we have the learning coefficient which we have given to the application by our choice and also there are three groups of weights which has total 24 weights for each of the eight attributes of the dataset. The input\_dataset multi dimensional array has 395 data records of the student performance based on the attributes described above. The values are in 1 or 0 format which represent yes or no. then we have the iterative counter which will help us to iterate the loop several times in order to train the data properly. First we find the h(equation 1) values and check which value is highest, only for that set we find the new updated weights. For finding the new updated weights we use standard competitive learning rule with normalization. The new updated weights can be used starting from the very next dataset of the situation of using them arises. This actions are performed for all the 395 datasets and we consider this to be as one iteration, repeating the same thing for the number of times we want to get the right weights and organizing the students into three different groups. After each dataset is evaluated and new weights are calculated we place the student into the respective group based on the group it falls according to the highest h(eq1) value as we calculate them by summing the multiple of input value and the weight for the group. After we process the datasets for each iteration we also print the new updated weights and the group to which the student belong to. There Is no specification to what each group represent but it is just separating them based on the dataset provided and weights. Decimal values are rounded upto 2 decimal places.

**Project Results:**

H(equation 1) output values can be represented as

h(eq1) values for Input data 1 is : 0.61 , 0.43 , 0.38

h(eq1) values for Input data 2 is : 0.66 , 0.29 , 0.5

h(eq1) values for Input data 3 is : 0.88 , 0.55 , 0.7

New Weights after 600 Iteration are :

0.02 , 0.02 , 0.02

0.06 , 0.07 , 0.06

0.06 , 0.1 , 0.1

After 600 Iteration grouping is:

Student 1 is in Group: 1

Student 2 is in Group: 1

Student 3 is in Group: 1

The output of the program looks something like this. Although it is not the entire output but just three values of each of the category of only last iteration. First we get the h(eq1) values by calculating them followed by updated weights and then we printout the iteration count the groups in which each of the students belong to. Each of the iteration shows us the weights being updated and the eq1 values being calculated and the group student belong to after that particular iteration of calculation.

**Conclusion:**

we have implemented one the good techniques in the neural network and have provided the system with proper dataset of student performance by taking them from UCI machine Learning repository. This way we can make preparations and plan steps to be taken through which students perform better next time.

**FollowUP:**

The program is almost complete as we have implemented all the features of simple competitive Learning. In the end of the iteration, we can see that the project is working and the students are organized into three different groups based on their data. Also the more the training the better the system can perform. So choosing the best of the initial weights can increase the result of the project.