****

**20CSIS01H**

**Data Mining & Warehousing**

**Group Number 1**

**Phase 2**

|  |  |  |
| --- | --- | --- |
| Mostafa Ebrahim Negm | SE | 153095 |
| Maryam Ashraf Diaa | SE | 162417 |
| Hussien fawzy | IS | 158267 |

**Support Vector Machine (SVM):**

Support vector machine is machine learning algorithm that classify problems of two groups. The goal of the SVM is to have best linear boundary so it can separate the n-dimensional spaces into classes so we can add new data points in its correct category. Hyperplane is the name of the best decision boundary. SVM create the hyperplane by choosing the most extreme points, and these extreme points are called support vector machine.

SVM has two types, which they are linear SVM and non linear SVM. The linear SVM is used for the data that is linear separable, which means the data can be classified in two classes by a straight line and this classifier called linear SVM classifier. The Non-linear SVM is used for the data that is non-linear separable, which means the dataset cannot be classified with straight line. The Non-linear classifier called SVM classifier.

How SVM works:

1. Hyperplane: hyperplane gets the best decision boundary. The dimensions of the hyperplane depend on the dataset features, if the dataset have 2 features then the hyperplane will be straight line but if the dataset have 3 features it will be 2-dimensional plane. The hyperplane should always have the maximum margin between the points.
2. Linear SVM: the linear SVM algorithm used to classify the data into two features by a straight line. The SVM helps to get the best line boundary and this boundary called hyperplane, so the SVM works by finding the closets points to the linear and it is called support vector. The SVM goal is to maximize the margin, and the margin is the distance between the vectors and the hyperplane. The optimal hyperplane is the hyperplane with maximum margin.
3. Non-linear: the data that cannot be separable with straight line we can classify it by adding one or more dimension. In Non-linear we add 3 dimensions not only 1 or 2 as the linear. So, the Non-linear calculates as (z=x2 +y2)

**Neural Network (NN):**

Neural network is simulation of the human neurons’ cells. It is made to connected many neurons with each other so the machine can learn, recognize pattern, and make a decision like humans. Neural networks work by learning by itself. Neural network consists of many neurons called units which they are arranged in series of layers, each of the layers are connected from both sides and some of them known as input layers. The input layers are designed to take many forms of information that needed to be processed. There is output units and it is the opposite side of the network and It signal how the neural network respond to the information. Between the input units and output units there is one or more hidden layers, those hidden layers are the brain of the neural network. The neural network is fully connected from the input units until the output units, each layer is connected with other layer by weight.

How NN works:

The neural network takes the information in two ways when the information is being trained and after trained. The information flows to the network through input units and then to the hidden layers and then turns the output to the output units. Each unit takes input from the unit on its left and then the inputs will be multiplied by the weight if the connection they are moving through. Each unit adds it weight to the input it receives, when the sum is more that threshold value the unit fire and trigger the unit that is connected to.

Artificial neural network is famous classification method, it collects the data and train the network on the data, so when input new data it can determine this input is close to which model.

**Dataset description:**

The dataset contains 14 attributes published by the Cleveland database. The data set refers to heart disease in patients. Where takes patient info and valued from 0 that means no presence to 4.

The attributes of the dataset contain:

* Patient Age as age
* Patient gender as sex
* Patient chest pain type as cp
* Patient resting blood pressure as tresrbps
* Patient serum cholestoral as chol
* Patient fasting blood sugar as fbs
* Patient resting electrocardiographic results as restecg
* Patient maximum heart rate achieved as thalach
* Patient exercise induced angina as exang
* Patient oldpeak as oldpeak
* Patient slope of the peak as slope
* Patient number of major colored by flourosopy as ca
* Patient thal as thal
* And target which 0 means healthy and 1 means have heart disease

**Compare between NN and SVM:**

**Accuracy:**

The NN accuracy was: 0.8250 which is 82.50%

The SVM accuracy was: 0.85 which is 85.24%

SVM accuracy: the dataset is 303 rows and 14 column and we train it on 242 rows and tested on 61 rows and we got accuracy 85.24%

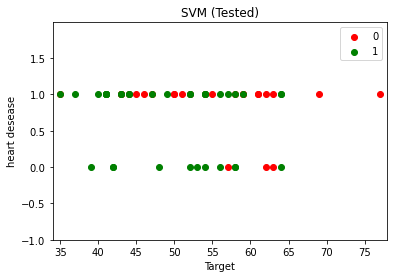
NN accuracy: the dataset is 303 rows and 14 column we train it on 200 row and tested on 103 rows and we got accuracy 82.50%

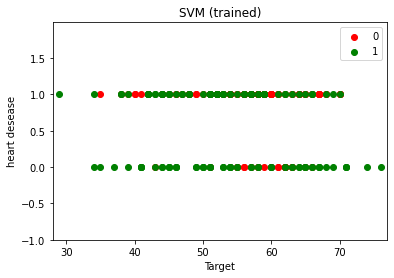
That means the both algorithms accuracy is in the same range over the dataset and the different between them is not that much. According to the SVM accuracy it better than NN, which means that SVM has slightly higher prediction accuracy than NN

**Visualization:**

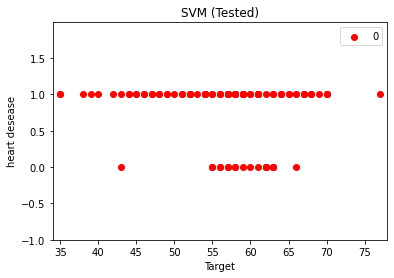
**SVM:**

**The data that have been tested:** here is the result of the tested data using SVM which 0 means they don’t have heart disease and 1 means they have heart disease.

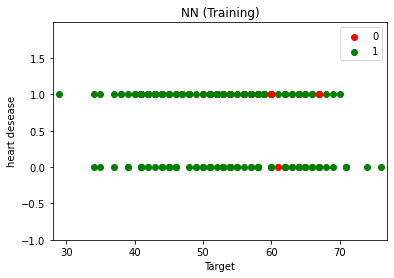
****

**The data that have been trained**: here is the result of the trained data where 0 means they are healthy and 1 means they have heart disease. ****

**NN:**

Here is the result of the tested data where 0 means no one have disease and 1 means the number of people who have heart disease

Here is the result of the trained data where 0 means no one have disease and 1 means the number of people who have heart disease



**Conclusion:**

The conclusion is the SVM is faster in training the data and gives accuracy slightly higher than the NN. In the other hand, the SVM only take small amount of dataset to train it despite the NN it can take as much as it can of data as input.

**References:**

[SVM Vs Neural Network | Baeldung on Computer Science](https://www.baeldung.com/cs/svm-vs-neural-network)

[How neural networks work - A simple introduction (explainthatstuff.com)](https://www.explainthatstuff.com/introduction-to-neural-networks.html)

[Support Vector Machine (SVM) Algorithm - Javatpoint](https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm)