CITY
UNIVERSITY OF LONDON
EST 1894

Student NO: 170052073

15/05/2022

INM427 – Neural Computing

By: Ajai Suresh

Lecturer: Dr. Kevin Ryan

Contents

1) Brief Description and Motivation of the Problem

Cardiovascular diseases are the number one cause of death worldwide. In 2019, an estimated 17.9 million deaths were reported from CVDs, accounting for 32% of all global deaths. 85 percent of these deaths were caused by a heart attack or a stroke. [5] Various data mining and neural network approaches have been used to determine the severity of heart disease in individuals.[3] In this project we will use two neural networks to predict whether someone has heart disease or not. The two neural networks are SVM (Support Vector Machine) and MLP (Multi-Layer Perception). The dataset used for this project was taken from Kaggle. [1]

2) Description of the Dataset Including Data Types

The dataset chosen to undergo this project is the largest dataset available so far for research purposes. There were 5 heart datasets combined into one dataset, where duplicate data were removed. The five datasets used for the combination are:

Cleveland: 303 observations

Hungarian: 294 observations

Switzerland: 123 observations

Long Beach VA: 200 observations

Stalog (Heart) Data Set: 270 observations

This adds up to 1190 observations, but 272 observations were duplicated, hence remaining with 918 observations for this project. This dataset has 12 columns, where there are 11 predictors and 1 target variable.

Looking at *table 1*, it can be seen that the dataset does not have any missing values, and looking further at *table 2*, it can be seen that there is slightly imbalanced data. As it is a small difference, there was no necessity to handle the missing values or the imbalanced data.

Attribute	Missing Values
Age	0
Sex	0
ChestPainType	0
RestingBP	0
Cholesterol	0
FastingBS	0
RestingECG	0
MaxHR	0
ExerciseAngina	0
Oldpeak	0
ST_Slope	0
HeartDisease	0

Table 1: Count of missing values

Heart Disease	Count
Positive [1]	508
Negative [0]	410

Table 2: Imbalanced Data

Furthermore, a boxplot was plotted to see if there were any outliers as shown in *figure 1*. The outliers were not removed as they could have an impact on the prediction of heart disease. Using label encoding, the categorical groups with strings were converted into numerical forms.

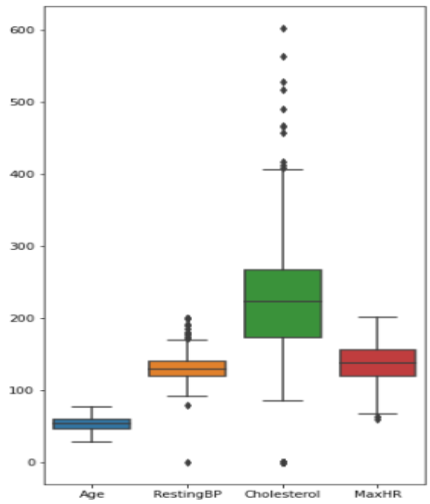

Figure 1: Outliers

The data was normalised between 0 and 1, as some of the columns had a different range of values compared to other columns. The target variable is represented by binary values 0 and 1, where 0 is negative and 1 is positive for heart disease.

3) Brief Summary of the Two Neural Networks

3.1) Support Vector Machine

SVM is a supervised machine learning algorithm that can be used for solving different classification problems such as binary classification and multi-class classification and regression analysis. [3]

3.1.1) Pros

- SVM works relatively well even with smaller datasets, as the algorithm does not rely on the entire data.
- SVM works more effectively for high-dimensional datasets. The reason behind this is that the complexity of the training data does not depend on the dimensionality of the dataset.
- SVM is suitable for both linearly and non-linearly separable data as we can use the kernel technique for nonlinearly separable data.
- SVM solution provides global minimum and not local minimum. This is shown by optimisation; ensures global optimality.
- SVM uses a subset of training data in the decision function, hence more memory efficient.

3.1.2) Cons

- The data points can be only put above and below the classifying hyperplane. This makes the SVM output directly the resultant classes and not the probability.
- Choosing an optimal kernel for SVM can often be difficult.
- If the classes are not separated properly, then SVM might not be the best option as it would not perform well.
- In addition to the kernel, it is difficult to fine-tune the hyper-parameters like C (soft margin cost) and gamma (Gaussian radial basis function) as it is difficult to visualise the impact of hyper-parameters.
- The more noise the data has, the less efficient is SVM.

3.2) Multi-Layer Perceptron

MLP is a class of an artificial neural network. The MLP consists of at least three layers of nodes: input layer, hidden layer and output layer. The hidden layer and the output layer will have nodes where it uses a nonlinear activation function. [4]

3.2.1) Pros

- MLP is suitable for complex non-linear problems.
- MLP works well with large input data.
- After training the data, the predictions are time efficient when using MLP.
- The accuracy can be achieved regardless of a small dataset or large dataset.

3.2.2) Cons

- MLP does not know to what extent each independent variable is affected by the dependent variable.
- The computations are difficult and are time-consuming.
- The quality of the training will result in the functioning of the model.

4) Hypothesis Statement

Looking into previous studies for both SVM and MLP; some hypotheses were made. SVM would relatively perform better and give a greater performance and has shown and given better results in the past compared with MLP. In addition, we can expect SVM to be more time efficient. However, previous studies also shows that the results should be improved by optimising the data. [6]

5) Description of Choice of Training and Evaluation Methodology

The choice of training was done by having 25% testing data and 75% training data. There were 688 observations for the training data. The classification performance was evaluated by three commonly used metrics; accuracy, recall and precision for both SVM and MLP.

Moreover, the training data was optimised by using grid search method along with 10-k fold cross-validation to improve the accuracy. For the optimisation the SVM used hyperparameters such as kernel, C (soft-margin-cost), and degree. Whereas for the MLP; there were a number of hidden layers, learning rate, and momentum. The result of this being a classification problem, the Softmax activation function was used at the output layer.

6) Choice of Parameters and Experimental Results 5%

As mentioned above, the MLP was built with an input layer, a hidden layer, and an output layer. Activation function ReLU was applied to the hidden layer and the SoftMax activation function was used for the output layer.

After doing the optimisation, the best hyperparameters for the SVM model were C = 1, degree = 2 and kernel = rbf. For the MLP model, the best parameters were lr = 0.3, module hidden size = 200 and optimiser momentum = 0.9.

7) Analysis and Critical Evaluation of Results 20%

Looking at *figure 2 and 3,* which represents the confusion matrices for both models. From the confusion matrices it can be seen that both models performed well, and that both models predicted positive for heart disease. *Figure 2 and 3* also shows that both of the models had similar accuracy, but SVM did perform better predicting positive cases for heart decision.

Furthermore, looking into table 3, it can be seen that SVM is giving better results at all parameters. SVM has a higher recall, precision and F1. This shows that SVM has performed better in detecting both positive and negative cases when compared with MLP.

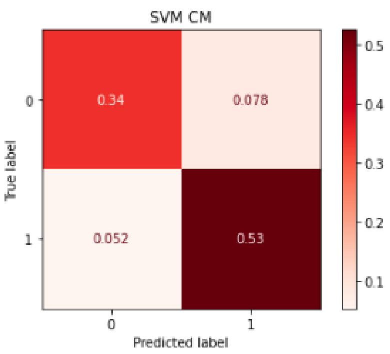

Figure 2: Confusion Matrix of SVM

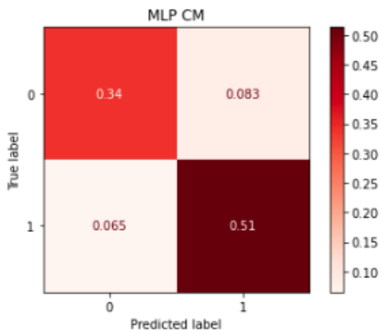

Figure 3: Confusion Matrix of MLP

	SVM	MLP
Precision	0.87 0.87	0.84 0.86
Recall	0.81 0.91	0.80 0.88
F1	0.84 0.89	0.82 0.87

Table 3: Precision, Recall and F1 of both SVM and MLP

Moreover, SVM has a better prediction and a higher accuracy as mentioned in the hypothesis section. SVM was also trained locally which means that less time was used on the SVM compared to the MLP, which required more time. However, there were only a small difference for between both the models.

On top of that, ROC curves and AUC of the models were obtained. Looking at figure 4 and 5; there are not a big difference in both of the models and the AUC for both of the models are the same, resulting 0.91.

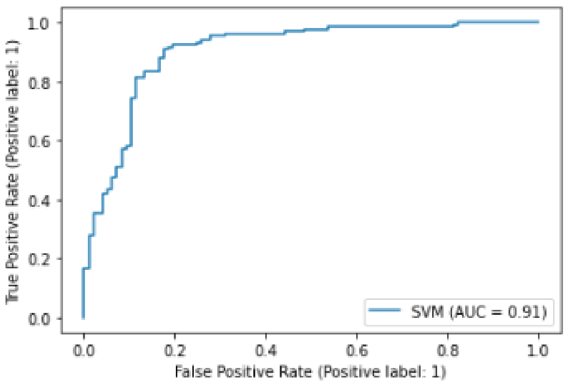

Figure 4: ROC curve of SVM

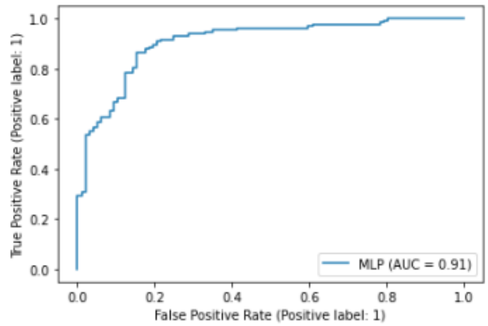

Figure 5: ROC curve of MLP

8) Conclusions, Lesson Learned, Reference, Future Work

To conclude, both models did give a high accuracy level when predicting heart disease. However, the results shows that SVM did perform better resulting that our hypothesis is true. It is also seen that, that tuning and optimising the parameters, will result to a higher accuracy in predicting.

There are a few lessons learnt after doing this project. One of the lessons learnt is that MLP will result a better accuracy if there are more hidden layers, but this also means that it will take more time. This was tested when I changed my hidden layers from 100 to 150, resulting a better accuracy but also more time.

For future work, other models can be used to see how they perform compared to SVM and MLP. Another factor that can be taken into consideration is that training data can be trained differently. A better data set cannot be found, as this is a dataset that has been combined from other datasets.

9) References

[1] Fedesoriano. (September 2021). Heart Failure Prediction Dataset. [Accessed: 15/05/2022]. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

[2] K, D. (2020). Top 4 advantages and disadvantages of Support Vector Machine or SVM. [Accessed: 15/05/2022] Medium. Available at: https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107 .

[3] Mohan, S., Thirumalai, C. and Srivastava, G. (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. IEEE Access, 7, pp.81542–81554. doi:10.1109/access.2019.2923707.

[4] Uniqtech (2019). *Multilayer Perceptron (MLP) vs Convolutional Neural Network in Deep Learning*. [Accessed: 15/05/2022] Data Science Bootcamp. Available at:

https://medium.com/data-science-bootcamp/multilayer-perceptron-mlp-vs-convolutional-neural-network-in-deep-learning-c890f487a8f1

[5] World Health Organization (2021). *Cardiovascular Diseases (CVDs)*. [Accessed: 15/05/2022] https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[6] Zanaty, E.A. (2012). Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification. *Egyptian Informatics Journal*, 13(3), pp.177–183. doi:10.1016/j.eij.2012.08.002.