Introductory Machine Learning Project

Individual Report 2A

Suresh Balaiyan-C3302159

Abstract. This report explains the basic application algorithms of artificial neural network(ANN) and Support vector machines(SVM) .For task-1 (a) ANN was used to classify the popular two spirals of Lang and Witbrock (1988) from their respective X,Y co-ordinates obtained from the given CSV file. For task-1(b) variations of the two spirals are generated using different configurations from the web-link- https://gist.github.com/45deg/e731d9e7f478de134def5668324c44c5 then the algorithm of the above task-1(a) was used to classify the generated variation of the two spiral. Task-1(c) discusses comparison of both SVM and ANN applied on task-1(a) and task-1(b). Next, task-2- Landsat Satellite Data Set is used to train and test an SVM classification model to predict the type of soil. Evaluation metrics such as confusion matrix and Jaccard index was used to measure the accuracy of the predictions and compare the performance of SVM model on both train and test data sets. Finally , for task-3(1)drug dataset with the historical data of patients, and their response to different medications was used to train and test a SVM and ANN model with kernel such as linear for SVM and 2 and 3 hidden layers with 50 and 200 units in each for ANN, illustrating that the patients with similar health record will respond to the drug type[3]. For task 3(b) - telecommunication dataset with demographic data of customers who are grouped into four service usage patterns was used to train and test SVM and ANN model , if demographic data can be used to predict group membership, the company can customize offers for individual prospective customers.[3]

Introduction: Machine learning applications are used in our everyday lives. The most popular applications are image recognition/classification, speech recognition, text recognition and many more. However, we are unaware of how and when these algorithms are used. In this report we discuss the well-established classification models such as Artificial neural networks(ANN) and support vector machines(SVM).

ANN are mathematical computations with the use of concepts of neurons/ perceptron's which is modelled after our own brain structure of neural networks. Activation functions such as sigmoid are used to transform the input value between values 0 and 1, which in turn is fed as inputs for the next layers of a feed forward neural networks. Using back propagation and chain rule and iterating the computations with the use of epochs, results to predict the expected output. [5]

SVM uses the kernel trick to transform the data up to infinite dimensions so the data can be separated and thus the algorithm is able to predict the new data with the same features. Hyperplanes with maximum margin separates the class labels. Hyperplanes are implemented in a n-1-dimension space. So, hyperplane for a 1 dimension is a point, for a 2 dimension is a line and for 3 dimensions is a plane and so on.[5]

In this report we use ANN to successfully classify two spirals and its variations, while SVM is used to predict a type of drug that would suit a patient and also predict the type of soil from spectral values and also SVM is used to classify telecommunication customers based on their demographic data's.

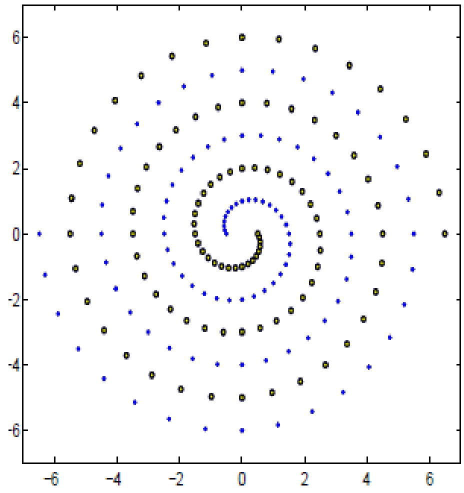
Table of contents

Q1. Variations of the Two-Spiral Task

1.a Two-Spiral task

Data Set Description:

Lang and Witbrock (1988) 194 training points is been used to train our ANN classifier to distinguish black and blue points on the spiral.



Approach on building the ANN model: This binary classier supervised ANN model is built based on the variations of the following combinations of hyper parameters such as:

1. Number of epochs(1000 to 200,000)
2. Number of hidden layers.(2 to 5)
3. Adjusting the weights (20 trainable neurons to 100)
   As we increase the value of the hyper parameters with various combinations of epochs, hidden layers and weights and bias , the error rate of the classifier model decreases to that of prediction values by implementing the gradient descent optimizer with a learning rate of 0.5.

The values of the hyper parameters are gradually increased as follows to obtain a well delimited two spiral:

```
Hidden layers: 2   Units per layer :40  epochs= 5,000 - 200,000
```

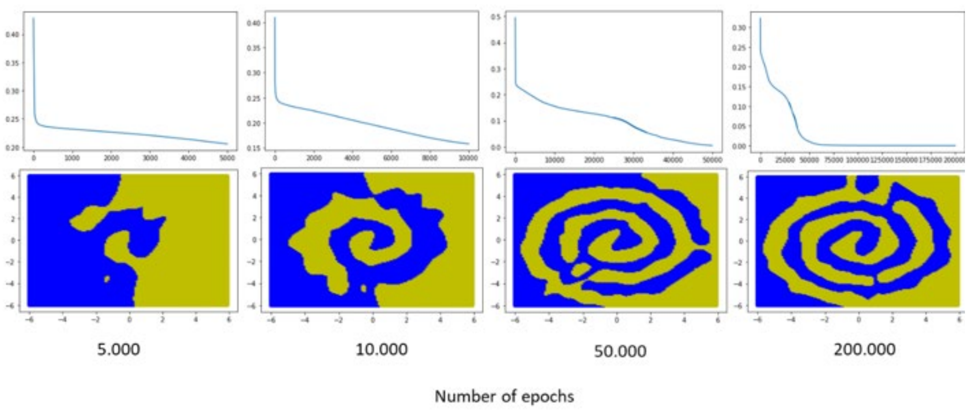


Figure 1-a-1 Shows the decreasing error rate and formation of delimited two spiral by varying epochs.

Hidden layers: 2 Units per layer :20 to 100epochs= 10,000-200,000

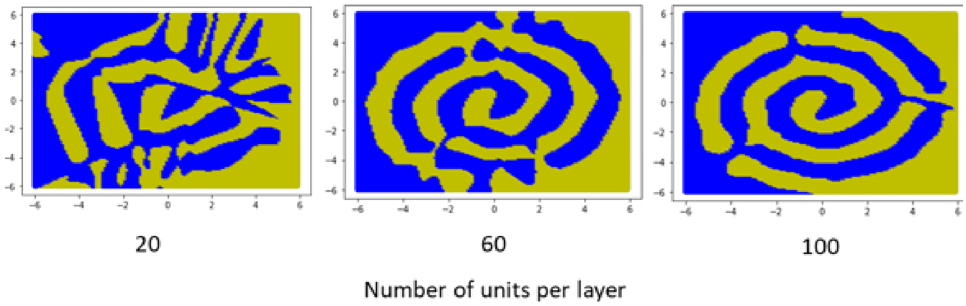


Figure 1-a-2 Shows the formation of delimited two spiral by varying number of units per layer.

Hidden layers: 3 Units per layer :40 -100 epochs= 10,000-200,000

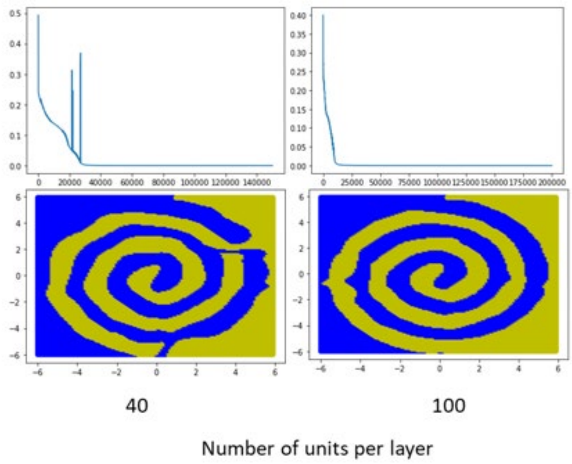


Figure 1-a-3 Shows the formation of delimited two spiral by varying number of units per layer.

The ANN model uses **back propagation and chain rule** to adjust the weights from the output layer to the hidden layers and to the input layer in order to minimize the error rate, so the model can predict the output that is expected . We use sigmoid activation function to obtain the probability of each layer which in turn used as input for the feed forward neural network: [2]

Conclusion: The ANN model begins to delimit the two spirals from 50,000 epochs and 100,000 epochs gives us a well separated spirals with 3 hidden layers that includes 100 units per layer, resulting zero error rate: [4]

1.b. Two spiral variation

Generating variations of the two spirals: Variations of the 2 spiral is possible by executing the following link code with variations of the parameters such as : Number of samples(N), density of spirals(D), number of spiral loops(L).

https://gist.github.com/45deg/e731d9e7f478de134def5668324c44c5

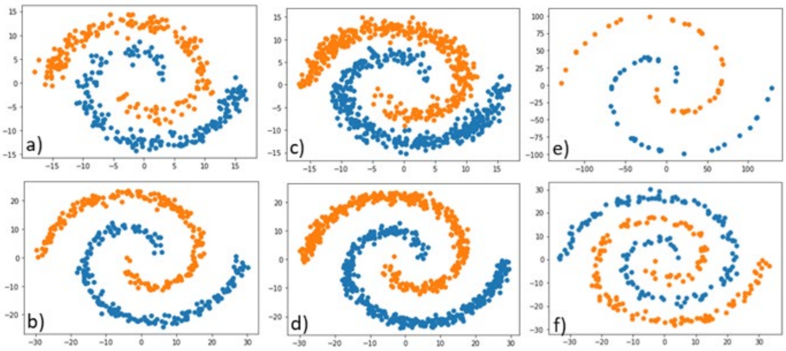


Figure 1-b – Two-spiral tasks generated by varying the parameters N, D and L. a) N= 200, D = 2/-2 and L = 2. b) N= 200, D = 4/-4 and L = 2. c a) N= 400, D = 2/-2 and L = 2. d) N= 400, D = 4/=4 and L = 2. e) N= 40, D = 20 and L = 2. f) N= 650, D = 2 and L = 3

Approaches on building the ANN model: We used the same algorithm as in the above a) but ANN classifier is unable to classify the two spirals and we obtain the following result for a spiral with :

`N=40, D=20 and L=2.`

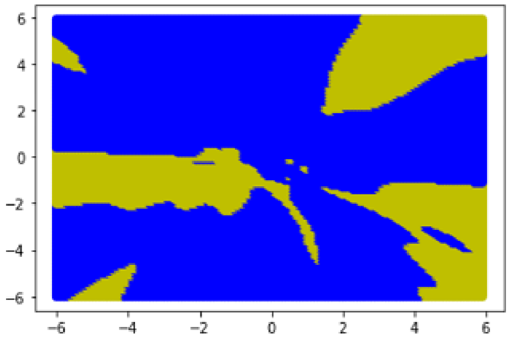


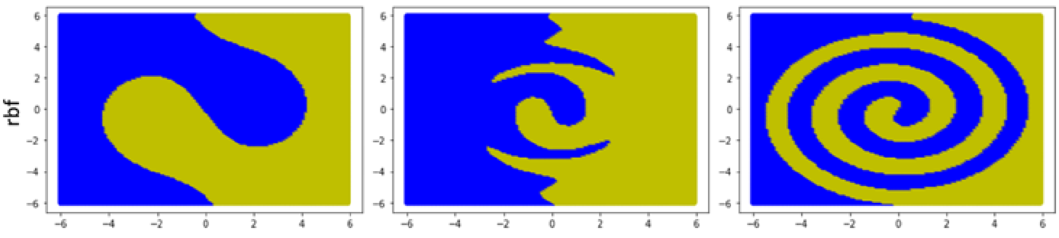
Figure 1-b2 – Results for a test with an ANN with 5 hidden layers with 200 units in each and 200.000 epochs for the two spiral task with N= 40, D = 20 and L = 2.

Conclusion: The ANN classifier failed to form a well delimited 2 spirals. Other classifiers such as SVM may give us a better result because SVM can transform the data into higher dimensional spaces(RBF-Kernel) and uses hyperplane to classify the data sets.

1.C. Compare ANN and SVM

1-a SVM vs ANN model:

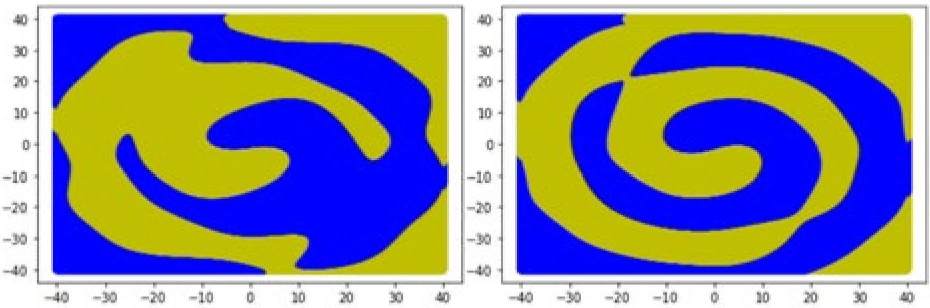
For Q1-a both the classification models such as ANN and SVM performed efficiently and produced a well delimited two spirals as follows.



(C=1, gamma=0.1) (C=1, gamma=0.5) (C=1, gamma=1)

Figure 1-c-1Results of the tests applying SVM with rbf kernel to the classification of the two spiral task of item 1-a.

While for the Q1-b only SVM model was able to classify the two spiral and predict the output correctly. This is due to the density of spirals and variations in the spirals generated.



C=0.08, kernel='rbf', gamma=0.01 C=0.2, kernel='rbf', gamma=0.01

Figure 1-c-2- Illustration of the calibration of the parameter "C" for the generated variations of the two spiral task

Conclusion: SVM performed efficiently on both the above tasks and resulted in a well delimited spiral . This is due to the SVM is able to be trained on datasets that are mapped non-Linear and kernel such as 'rbf' is more robust to classify the data linearly ,while ANN failed to classify the Q1-b task because ANN uses connection layers and activation functions such as sigmoid functions to classify non- linear data, which was not efficient enough to delimit the Q1-b task.[5]

Q2. Statlog (Landsat Satellite) Data Set

Data Set Description: The train data set given was used to train an SVM classifier model . The data set represents spectral bands . The dimensions of the train data set are : 4434 rows * 37 columns. The last column represents one of the given classes such as class[1,2,3,4,5,7].

Th test data set consists of 2000 rows* 37 columns. Again, the last column represents one of the classes such as [1,2,3,4,5,7].

Kernel and hyper parameters: Based upon grid search to find the best parameters to train the SVM classifier , we obtain the following as the best parameters :

```
Best parameters:{'C':1,'gamma':0.0001,'Kernel': 'rbf'}
```

Where **C=1** defines the misclassification that the model can allow during classifying the classes. As C value is high the model does not allow misclassification of class labels , while on the other side high C value can lead to overfitting the classifier model. In our case the C value is not of high value, so we do not expect our model to overfit ie training the model on possible noise/outliers in the train datasets.[2]

Gamma hyper parameter 0.**0001** defines the influence of the data values on other data values during classification , thus lower gamma value in our case has higher influence of the observations nearby. In other words, the observations in the data set is like weighted neighbours. Thus 0. 0001 defines higher importance of datasets that are close to each other in building the classification model. [2]

RBF Kernel is based on the squared **Euclidean** distance of the train datasets with the product of the gamma value chosen:

$$K(x, x') = \exp(-\gamma||x - x'||x^2) \quad (1)$$

Thus, RBF Kernel uses kernel trick to classify as a linear classifier to distinguish the class labels[1,2,3,4,5,7] by the illusion of transferring the data points to higher dimensions to form hyperplanes between the classes: [1]

Train the classifier model: As a result of fitting the train dataset using the hyper parameters described above, we were able to achieve best of the classifier in predicting the Y or class labels on the train datasets as follows:

```
Model accuracy: 91.32 %
```

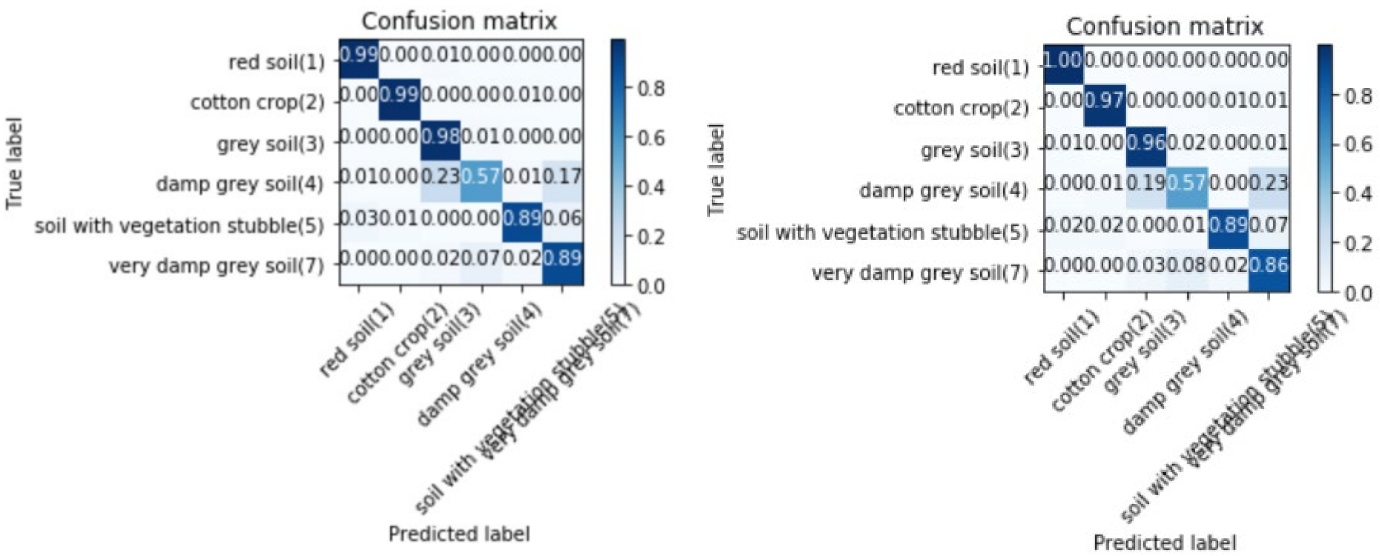
Test the classifier model: As the classifier has given high accuracy rate on train data set, we proceed to test on the test dataset. Because we are testing the classier on a **different** data set rather than on the train dataset, we can assure the validity of the model and the accuracy on the test dataset in predicting the output class labels, which is **90%** :

```
SVM accuracy on test data set: 90 %
```

Evaluation metrics:

1. Confusion matrix:

Train Data set **Test Data set**



2.

Figure 2-1 Shows the normalized confusion matrix with the scores of prediction accuracy on each class labels of train data set(on left) and test dataset (on right)

From the above matrix we can see that the SVM model performed efficiently on both the train and test dataset. Except for class labels 4,5 there were minor mis match in the predictions.

Jaccard similarity score : This similarity score indicates the percentage of predicted class labels to that of the true labels or actual values in the test data set[3], which is 90%(same as the above confusion matrix) and 1795 out of 2000 test data records.

```
Jaccard similarity accuracy score on test data set: 90 %
```

Out of 2000 test data samples , our SVM model correctly predicted 1795 data records.

```
Jaccard similarity accuracy score out of 2000 data records: 1795
```

Conclusion: The SVM model has performed well on both the train and test data sets with maximum accuracy rate 91% and 90% respectively, this shows the trained SVM model is not overfitted nor underfitted and uses robust kernel such as 'rbf' to classify using metric squared Euclidian distance to predict accurately , the non-linear soil type class labels.[6]

Q3-1 Comparison ANN and SVM classification

Drug Data set:

Data Set Description: Given historical data of patient's records, and their response to different medications, with the dimensions of 200 rows * 5 columns : [3]. By using this data set we can predict a new patient with a given health metrics to respond to one of the five medications.

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY

Figure 3-1-1 Shows the first five records of the drug data set

Data set pre-Processing: Non -numeric values of the data set has been transferred to numeric values as follows:

```
Sex :[ Male, Female]{0,1}
```

```
Blood Pressure : ['LOW', 'NORMAL', 'HIGH'][1,2,0]

Cholesterol :[ 'NORMAL', 'HIGH'][1,0]

Drug:[ A,B,C,X,Y][1,2,3,4,5]
```

A snapshot of the dataset after numeric conversions is as follows:

```
array([[23, 0, 0, 0, 25.355],
       [47, 1, 1, 0, 13.093],
       [47, 1, 1, 0, 10.113999999999999],
       [28, 0, 2, 0, 7.797999999999999],
       [61, 0, 1, 0, 18.043]], dtype=object)
```

Split the data set into train and test data set: We split the above-Mentioned data set into train (80%) and test(20%) data set.

Upon grid search , we find the following as the best hypermeters and the kernel to fit and train the train data set:

Train the SVM model on the train data set: `Best parameters: {'C': 1, 'gamma': 0.0001,'Kernel': 'linear'}`

Train the ANN model on the train data set: Different configurations with different activation functions, 2 and 3 hidden layers and between 50 and 200 units in each.

After training the classifications ANN and SVM models – we tested on the test dataset.

Evaluation metrics: SVM

Training Dataset **Test Dataset**

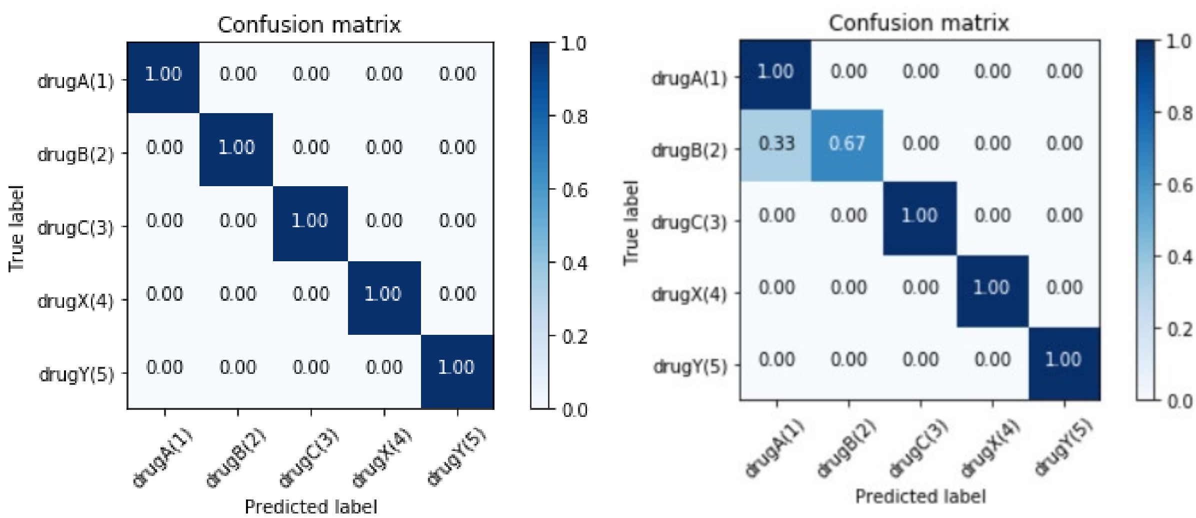
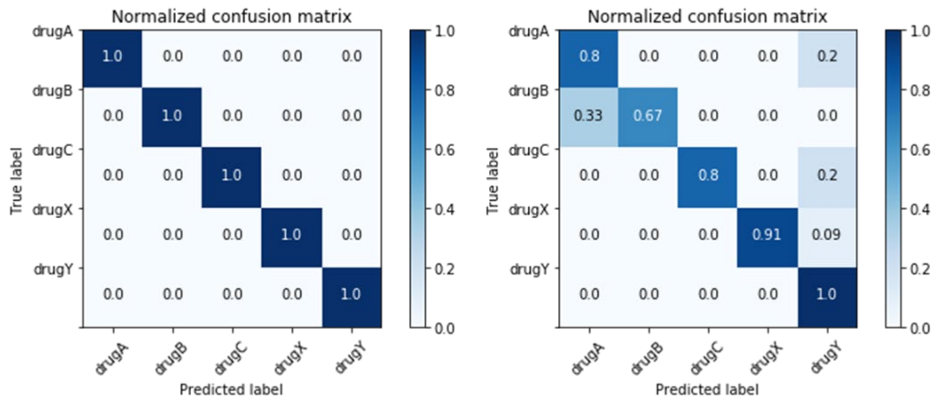


Figure 3-1-2 Shows the normalized confusion matrix of train data set(on left) and test data set (on right) for SVM

Evaluation metrics: ANN



Training dataset **Test dataset**

Figure 3-1-3 Shows the normalized confusion matrix of train data set(on left) and test data set (on right) for ANN

Conclusion: The SVM model and ANN performed efficiently both on train and test dataset with similar train and test results.

Q3-2 Comparison ANN and SVM classification

Telecommunication Dataset

A telecommunications provider has segmented its customer base by the service usage patterns, categorizing the customers into four groups: [3]. If demographic data can be used to predict group membership, the company can customize offers for individual prospective customers. It is a classification problem. The target field, called “custcat” has four possible values that correspond to the four customer groups, as follows:

1- Basic Service

2- E-Service

3- Plus Service

4- Total Service

A snapshot of the data set:

	region	tenure	age	marital	address	income	ed	employ	retire	gender	reside	custcat
0	2	13	44	1	9	64.0	4	5	0.0	0	2	1
1	3	11	33	1	7	136.0	5	5	0.0	0	6	4
2	3	68	52	1	24	116.0	1	29	0.0	1	2	3
3	2	33	33	0	12	33.0	2	0	0.0	1	1	1
4	2	23	30	1	9	30.0	1	2	0.0	0	4	3

Figure 3-2-1 Shows the snapshot of the data set

Data set pre- Processing: Transform the data set into its equivalent standard value (except custcat)by using its mean and standard deviation value:

```
X=preprocessing.StandardScaler().fit(X).transform(X.astype(float))
```

```
X:[0:5]
```

```
array([[-0.02696767, -1.055125  ,  0.18450456,  1.0100505 , -0.25303431,
        -0.12650641,  1.0877526 , -0.5941226 , -0.22207644, -1.03459817,
        -0.23065004],
       [ 1.19883553, -1.14880563, -0.69181243,  1.0100505 , -0.4514148 ,
         0.54644972,  1.9062271 , -0.5941226 , -0.22207644, -1.03459817,
         2.55666158],
       [ 1.19883553,  1.52109247,  0.82182601,  1.0100505 ,  1.23481934,
         0.35951747, -1.36767088,  1.78752803, -0.22207644,  0.96655883,
        -0.23065004],
       [-0.02696767, -0.11831864, -0.69181243, -0.9900495 ,  0.04453642,
        -0.41625141, -0.54919639, -1.09029981, -0.22207644,  0.96655883,
        -0.92747794],
       [-0.02696767, -0.58672182, -0.93080797,  1.0100505 , -0.25303431,
        -0.44429125, -1.36767088, -0.89182893, -0.22207644, -1.03459817,
         1.16300577]])
```

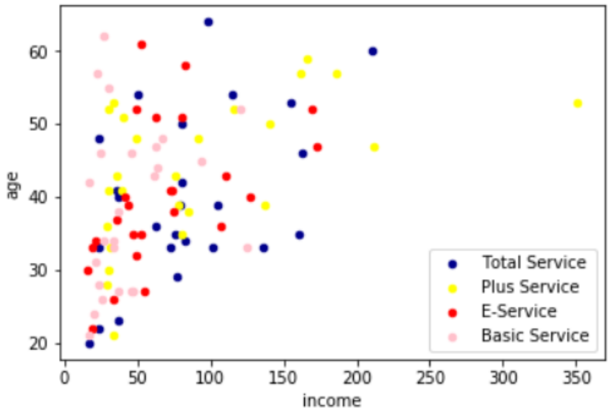


Figure 3-2-2 Shows the transformed standard scalar value of the data set and the scatterplot with the distribution of service types to that features -Age and income.

Split the data set into train and test data set:

We now split the data set into train(80%) and test(20%) data set and based on grid search on the train dataset we obtained as the optimum parameter to train the SVM model.

Train the SVM model:

```
Best parameters: {'C': 0.01, 'gamma': 0.0001,'Kernel': 'linear'}
```

Train the ANN model:

we implemented the same code of the last model with some minor changes to handle the new data set.

Evaluation metrics: SVM

Training dataset **Testing dataset**

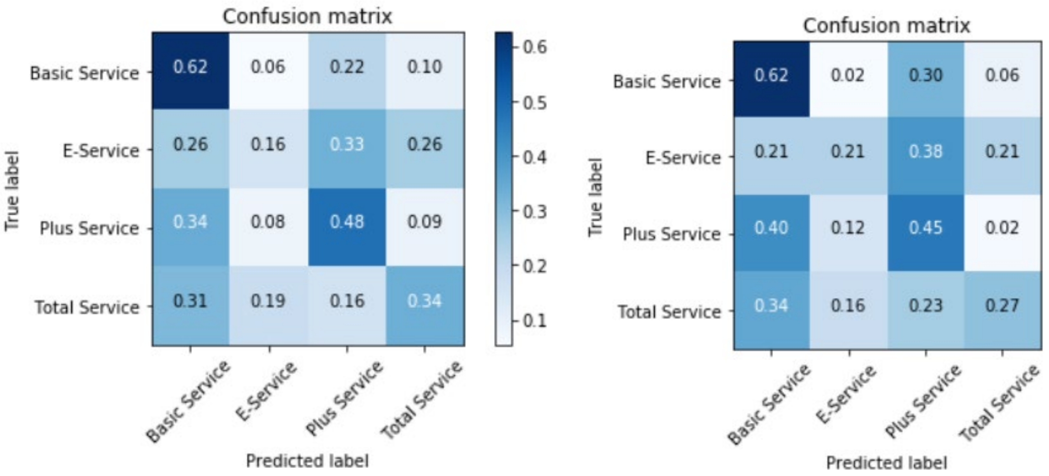
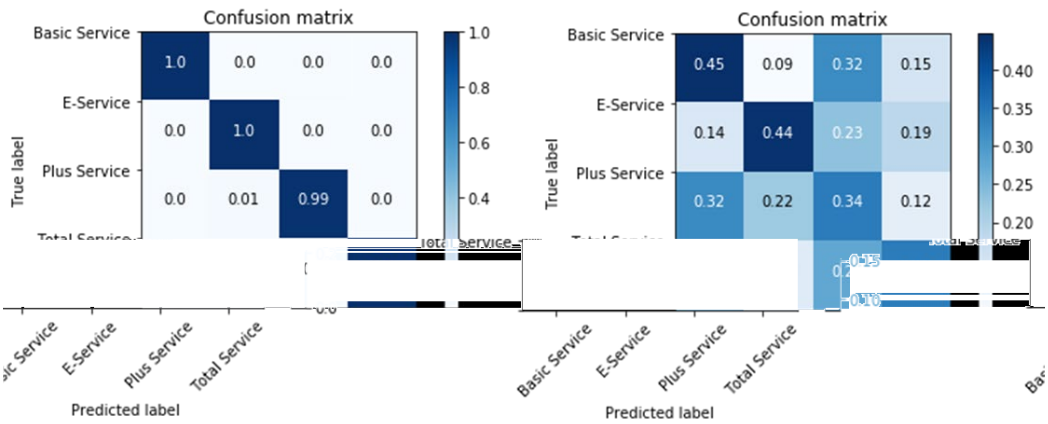


Figure 3-3-3 Shows the normalized confusion matrix of train data set(on left) and test data set (on right)

Evaluation metrics: ANN



Training dataset **Test dataset**

Figure 3-3-4 Shows the normalized confusion matrix of train data set(on left) and test data set (on right)

Conclusion:

Both ANN and SVM models classified with similar test results with the type of service-Approximately 40% . This is likely due to the existence of outliers in the dataset values that make the SVM and ANN classifier overfit and fail to classify accurately. Thus, overfitting enables better training results but not with test datasets. Removing outliers/noise in the dataset may give us high prediction accuracy. Training and testing on different data sets is recommendable in this scenario to improve prediction accuracy rate.[7]

References

1. Radial basis function kernel. (2020). Retrieved 23 April 2020, from https://en.wikipedia.org/wiki/Radial_basis_function_kernel
2. Neural network part 3 thanks — 3Blue1Brown. (2020). Retrieved 27 April 2020, from https://www.3blue1brown.com/nn3-thanks
3. IBM Developer Skills Network - Labs. (2020). Retrieved 16 April 2020, from https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/coursera/ML0101EN/ML0101EN-Clas-SVM-cancer-py-v1.ipynb
4. The Scuffle Between Two Algorithms -Neural Network vs. Support Vector Machine. (2020). Retrieved 28 April 2020, from https://medium.com/analytics-vidhya/the-scuffle-between-two-algorithms-neural-network-vs-support-vector-machine-16abe0eb4181
5. ANON (2020). Retrieved 4 May 2020, from https://strathprints.strath.ac.uk/48403/1/ANN_SVM_v4.pdf
6. An in-depth guide to supervised machine learning classification. (2020). Retrieved 27 March 2020, from https://builtin.com/data-science/supervised-machine-learning-classification
7. Machines, 2., & Gupta, A. (2020). 25 Questions to test a data scientist on Support Vector Machines. Retrieved 9 May 2020, from https://www.analyticsvidhya.com/blog/2017/10/svm-skilltest/