Thanadon Fuengworatham 29392802 tf2n17@soton.ac.uk

Machine Learning Assignment Report

1 Neural Network Approximation

The main propose of this section is to show the ability of ANN algorithm as function approximators in classification problems and compare the results with Bayes theory. Follows are the steps to prepare the experiment of this section.

1. Generating data from 2 classes of Gaussian distribution with

$$m_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, C_1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, m_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, C_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

2. Drawing the boundary line by contour function according to the Bayesians posterior equation of Class1 as presented in the following equation: -

$$P(w_1|x) = \frac{1}{1 + \exp(\frac{-1}{2}(x - m_2)^t C_2^{-1}(x - m_2) - (x - m_1)^t C_1^{-1}(x - m_1))} \tag{1}$$

Calculating the posterior probability of all points on the plane in Figure 1, then drawing the contour line at where posterior probability is equal to 0.5. This will give the Baye's boundary as the black dashed line on Figure 2.

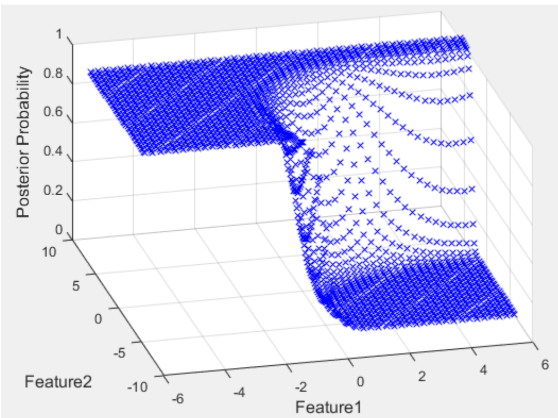

Figure 1: 3d plot of posterior probability of class1 to all points on the plane

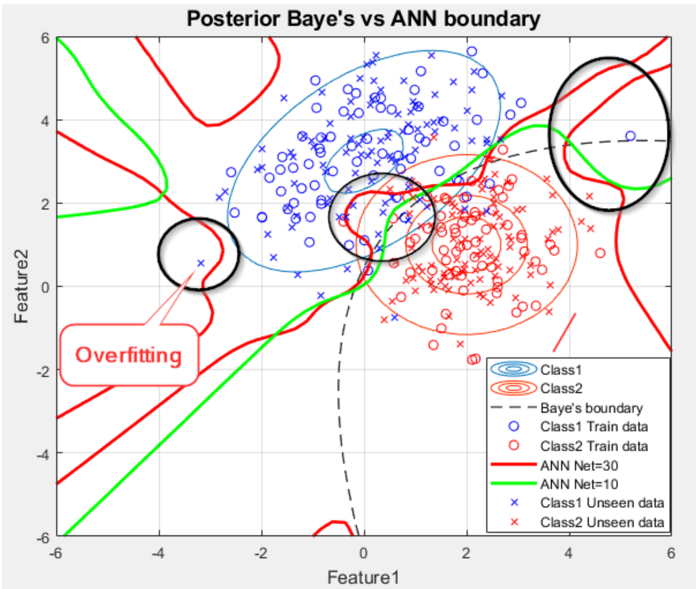

Figure 2: Two-class data separated by Bayes' and ANN boundary with net size 10 and 30

3. Build classifier model from ANN by preparing the data as follows: -

- Constructing feature matrix 200x2 by concatenating 100 sample data from each class
- Construct target matrix 200x1 with class label of 1 and 0 for class1 and 2 respectively.
- Training two models with different net sizes of 10 and 30 and applying them to every coordinate on the plane then plotting the contours where results equal to 0.5 on top the data in Figure 1.

Comparison between Baye's and ANN with different net size

According to Figure 2, it can be seen that the ANN boundaries are very good at classifying seen data. This can be noticed from the top-right circle of the figure where the ANN boundaries can perfectly classify the data point, while it is ambiguous for Bayes boundary. However, the ANN boundaries are sensitive with the unseen data (X point) especially the ANN with net size 30, which is **over-fitting** the data as shown at the bottom-left of the figure. While the Bayes boundary is **more generalized** with the data.

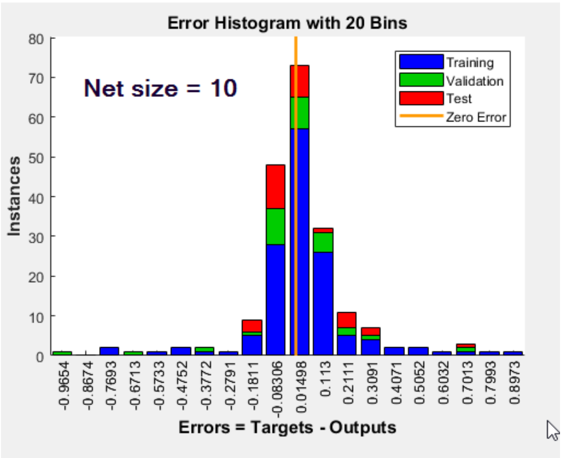

Figure 3: Error of testing result where net size = 10

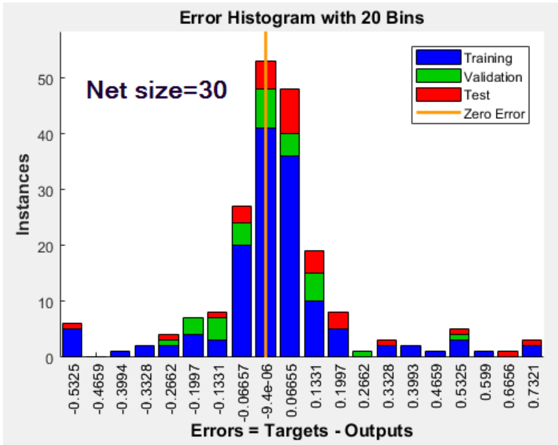

Figure 4: Error of testing result where net size = 30

Considering the different net sizes of ANN, the bigger net size of 30 produced more **over-fitting** results as shown in the circle at the center of the figure, where the boundary with 30 nets (red) misclassifies a lot of unseen data (X in blue) at the overlapping zone. According to Figure 3 and 4, net size 10 produced less error than net size 30. This can be seen from most errors distributed around zero error line for net size = 10. Thus, the proper size of net must be carefully selected to suite the condition.

Choosing proper net size in hidden layers

Too small net size can cause under-fitting problem and too big net size leads to over-fitting problems. Thus, to produce the best result with ANN algorithm, proper number of networks unit in hidden layer should be carefully selected. There is no fixed rule to choose the best number; however, there are some factors such as number of data, using algorithm, type of activation function, etc.[1] should be considered. Most importantly, performance measurements with different network size are required to choose the best parameter values according to the test results.

Conclusion

Baye's boundary can be used when there is limited number of sample data due to its generalization. However, for large data set, ANN is a better choice to produce higher accuracy. Nevertheless, ANN has advantage over Baye's that it can deal with both limited and large dataset by tuning the net size and the over-fitting problem can be solved by selecting the proper size of the networks in hidden layer.

2 Mackey-Glass Time Series

This section is to use regression model to predict the Chaotic time series data generated from Mackey-Glass model. The main purpose is to compare performance between Linear regression and ANN in 2 types of prediction, one-step-ahead and free-running mode.

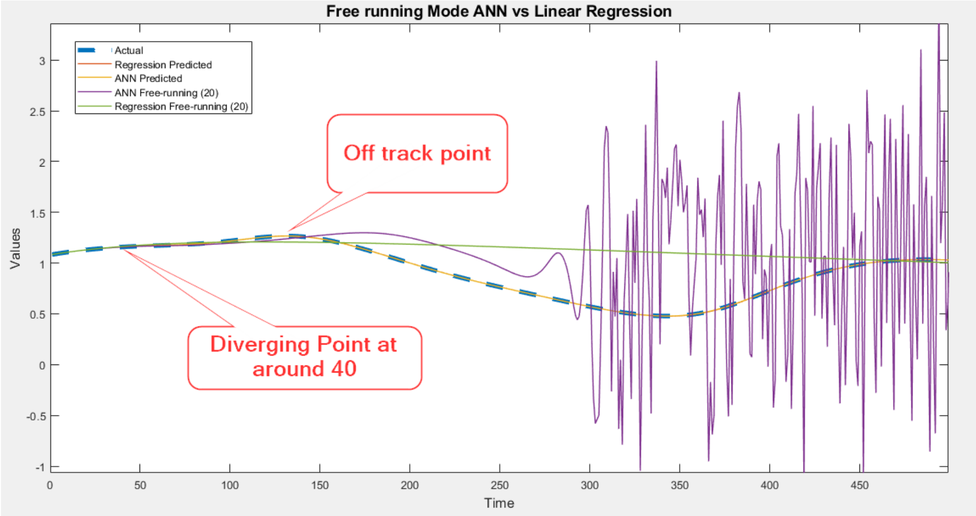

Figure 5: Comparison of predicted results of test dataset from Linear regression and ANN in one-step-ahead and free-running prediction

One-step-ahead Linear Regression vs ANN

According to Figure 5, both can produce perfect predicted results in which they follow the actual line over the time series. The accuracy plot on Figure 6 and 7 show the actual values and predicted values go along perfectly on the linear line.

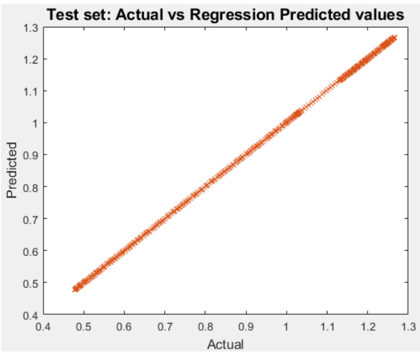

Figure 6: Accuracy plot comparing between actual and predicted results by Linear regression

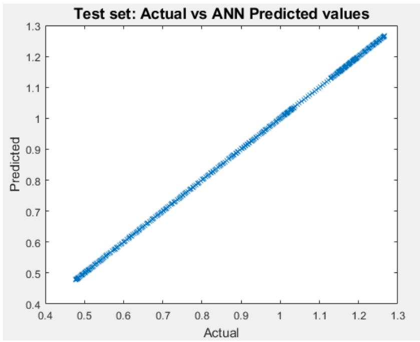

Figure 7: Accuracy plot comparing between actual and predicted results by ANN

Free-running Mode

According to Figure 5, there are three important phases considered on the free running line of both Linear regression and ANN predictors

1. **Diverging point** the lines show perfectly result until at around time 40, the lines started diverging from the actual values. This can be presumed that when the model completed one cycle of 20 series at time 40, where there is no actual data from the first phase (1-20) left in the input series, both Linear regression and ANN predictors will start to produce some errors. Therefore, it can be assumed that larger series (p) can maintain the predicted values on the actual track. This assumption leads to the experiment with different p series in the Table 1.
2. **Off-Track point** according to Figure 5, at around time 150, the accuracy from Regression and ANN predictors dropped rapidly. This made both of the predictors off from the track of actual values.
3. **After off-track** Linear Regression line produces stable trend lines until the end. Interestingly, the results from ANN ran steadily for a while then turned into the oscillations. These oscillations are sustained by itself over the time until the time series ended. This can be presumed that the ANN algorithm has ability to memorize oscillations from the Chaotic time series data which have been learnt from the training data at the training phase.

Comparison ANN results from different p series

P SERIES	DIVERGING POINT	OFF-TRACK POINT	ARE OSCILLATIONS SUSTAINED?
10	40	80	Yes
20	40	150	Yes
40	40	80	Yes

Table 1: Diverging and Off-track point on different time series size (p) in test dataset

This part is motivated from the assumption that larger p should produce free-running mode being on the track longer. However, from the results in the Table 1, it shows that increase of p to 40 does not help last the on-track point longer as well as the decrease of p size to 10. According to Figure 8, all predictors start diverging at time 40 (Point 1) and being off-track at around time 80 (Point 2) except p=20 that is able to be on track almost twice longer of the others (Point 3). In addition, all p sizes have ability to sustain the oscillations after being off the track.

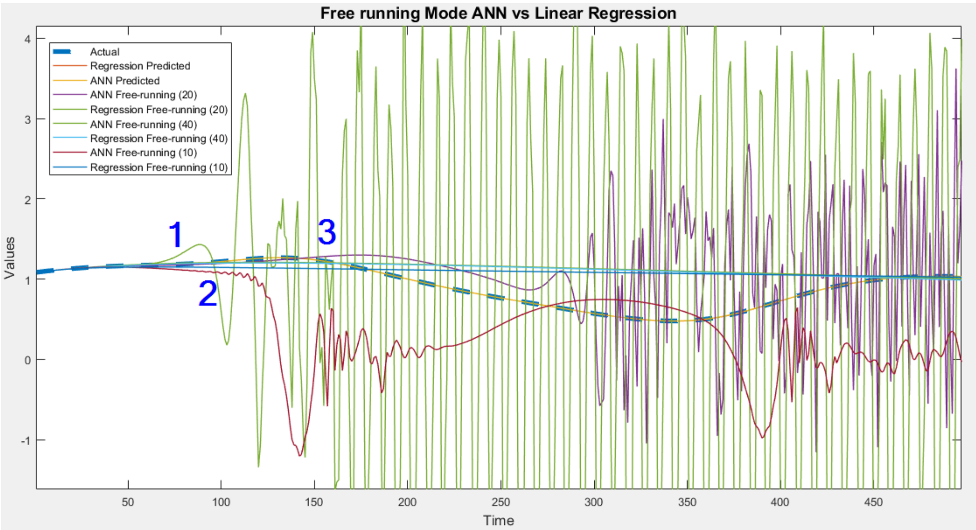

Figure 8: Diverging and Off-track point on different time series size (p) in test set

Conclusion

One-step-ahead prediction by Linear Regression and ANN model work perfectly on the Mackey-Glass data model. While, for free-running mode prediction, both predictors can maintain the accurate results for a period of time before breaking at the off-track point. Attempt to tune the time series size could not improve performance of both Linear regression and ANN predictors. However, ANN predictor has the ability to sustain the oscillations over the period in any time series size.

3 Financial Time Series

This section is to apply Linear regression and ANN models to predict the FTSE100 future stock price from 5-year history data, 3 Dec 2012 to 2017. The data is partitioned into two sets, first 2/3 (3 Dec 2012 5 Sep 2016) as training set and another 1/3 (6 Sep 2016 1 Dec 2017) as test set. First part of the section is to present the comparison between predicted result from only history prices and the result from both history prices and volumes. Second part will look at opportunities to make money in the stock market from these predictors.

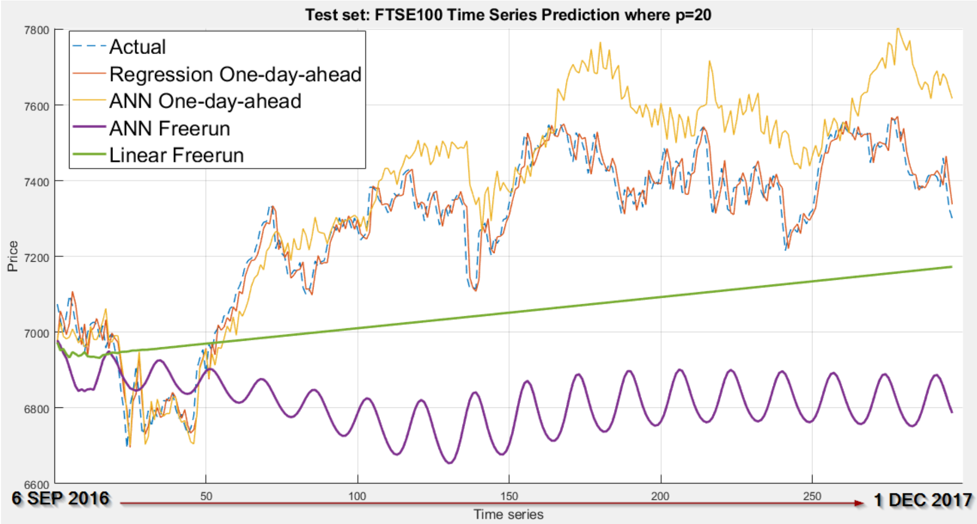

Figure 9: Price Prediction of test dataset from history prices comparing between the 4 of the predictors

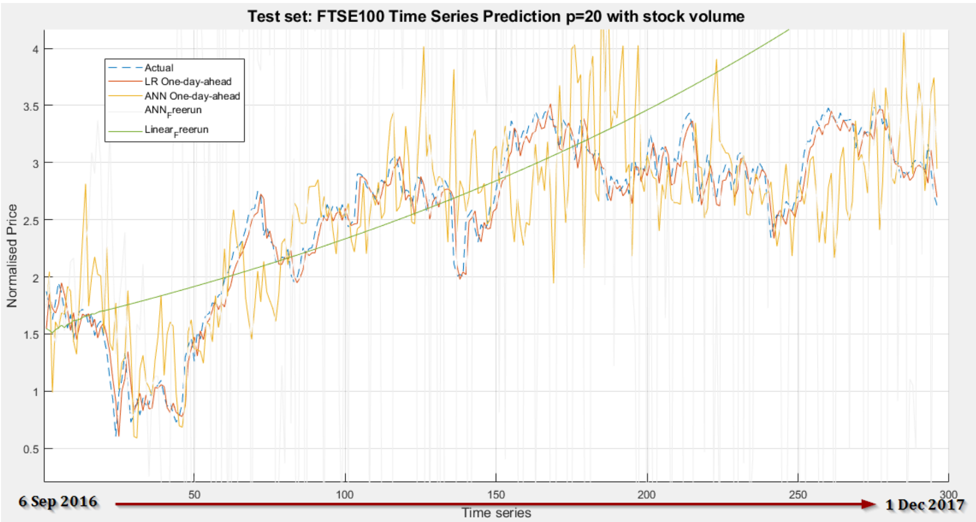

Figure 10: Price Prediction of test dataset from history prices and volumes, comparing between 4 of the predictors

Prediction with history prices

Figure 9 shows the prediction results using history prices. It is apparent that Linear regression predictor can produce more accurate results than ANN in one-day-ahead prediction. The regression predictor can maintain the accuracy over the period, while ANN shows high error rate from around middle of the period. According to Figure 11 and 12, it is clear that regression predictor performs better than ANN. The regression line goes along with the actual prices, while ANN results diverge from the accuracy line. Additionally, with regards to free-running prediction, both regression and ANN show inaccurate result where both are off the track since the starting point. However, there are still some hopes on free-running model that if we could maintain the accuracy more than one day ahead, it would be a more powerful tool for making money from the stock market.

Prediction with history prices and volumes

To improve the performance of the prediction, history volumes are included as additional features. Thus, there are 40 features in total, 20 from price series and another 20 from volume series. In free-running mode, instead of calculating only next days price to be fed back as a new input over time series, next days volume is also additionally calculated to maintain the volume series.

Figure 11: LR 1-day predicted result

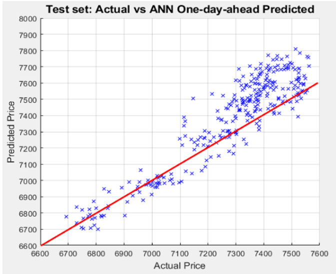

Figure 12: ANN 1-day predicted result

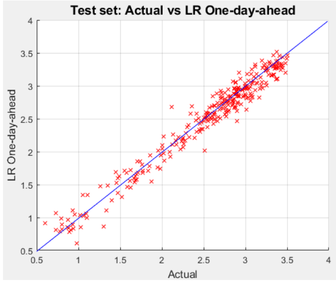

Figure 13: LR 1-day predicted with volume

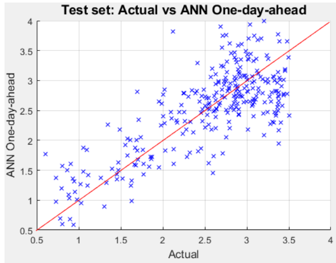

Figure 14: ANN 1-day predicted with volumes

According to Figure 10, overall, using volumes as new features makes more noises in the predicted results because the result relies on 2 parameters making the result not as stable as before. According to Figure 14, the noises affect the result of one-day-ahead prediction of ANN where the predicted results are sparser compared to Figure 12. However, this change improves ANN free-running predictor where the purple line in Figure 10, it can follow the actual-price line at the first short period, while in Figure 9 the line was off the track since the starting point. This can be presumed that using volumes as features can improve the performance of ANN models.

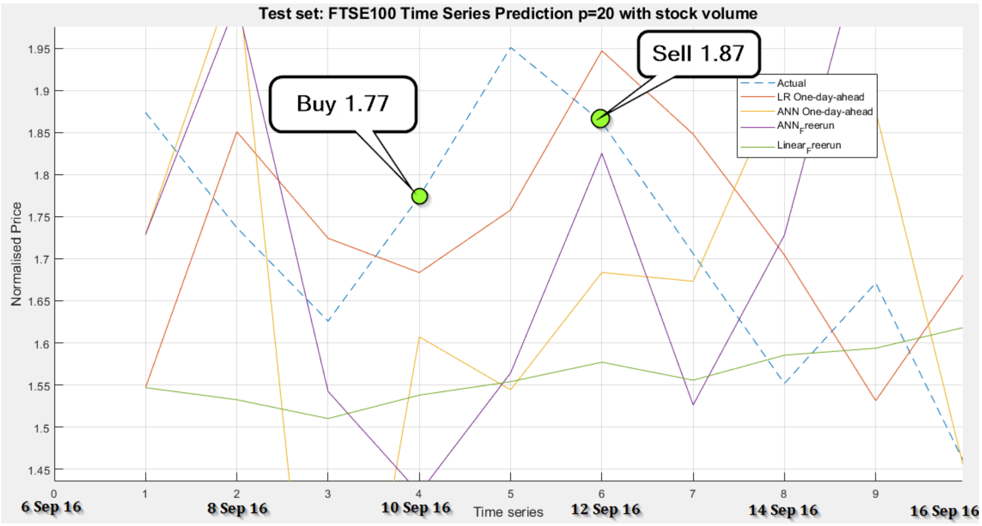

Figure 15: Predicted prices of test dataset in 10-day period from 7 to 16 September 2016 with buy and sell signal from ANN free-running predictor

Opportunities to make money

According to Figure 15, the tomorrows prices predicted by one-day-ahead predictors are one-day lagging behind the actual price. This means the predicted prices are actually just today prices, not tomorrows. As a result, the one-day-ahead predictors cannot be used as good indicators to make profits in stock market.

On the other hand, even though ANN free-running predictor shows the same lagging results, it can predict the prices in the longer future. Thus, if the trend line was correct, it would be possible for us to make profits by following this trend line. According to Figure 15, the purple line starts by following the actual line until it breaks off from the actual trend on 13 Sep 16. This can be assumed that the ANN free-running predictors can predict reliable trend within 7-day-ahead period. With this assumption, if we were to buy the stock at the lowest predicted price on 10 Sep 16 and sell it on 12 Sep 16 as shown in Figure 15, we would make the profit for 0.1 point of normalized price. Thus, the ANN free-running model has potential to make money in the stock market.

References

[1] FAQ in comp.ai.neural-nets, Part 3 of 7: Generalization Section - How many hidden units should I use?,
`http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-10.html`