Neural Network Model Case Study

Shengyu Wu
UNSW
MATH 5836
Z5317718

I. Introduction

Neural networks along with deep learning models has led to very good performance on a variety of problems, such as improved guidance systems, development of power trains, condition monitoring. Critical tasks of neural networks include classification, prediction, clustering, associating.

However, there are some challenges too. For instance, while developers need more control over the details, the duration of network development can be time-consuming, and the model is also computationally expensive. Note that optimization for deep networks is currently updated. In stochastic gradient descent (SGD), the weights are based on training example instead of as a whole in gradient descent to minimize loss faster. But it is also noisier.

To assess the profitability and the sustainability impact of harvesting the abalone, we are interested in classifying and predicting the age of a particular abalone. In this case, A deep learning model is helpful to analysis the dataset.

This paper investigates the main findings of using neural network in the task of classification problem. The major goal is to evaluate the best model to classify abalones into 4 classes base on ring age. Data were collected about a large sample of abalone and can be found in abalone.data.

The rest of the paper is organised as follows. First, we investigate the optimal hidden neurons for a single hidden layer. Second, select the optimal learning rate using the optimal hidden neurons. Third, select the appropriate number of hidden layers with the optimal hidden neurons. Then compare training and test performance of Adam and SGD algorithms to evaluate the best model using a confusion matrix. Finally, we conclude the paper with a discussion for future research.

II. Methodology

A. Data Processing

Researchers collected the following information about each abalone:

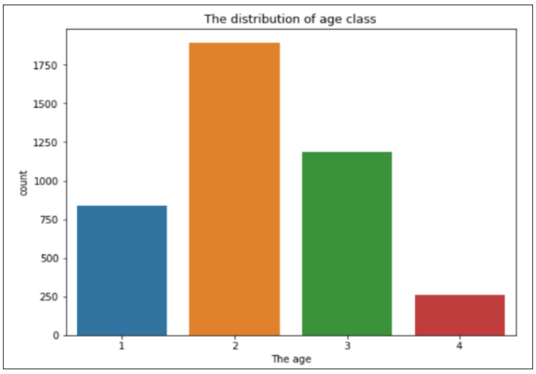
Sex - Male, Female, or Infant
Length - (mm) longest shell measurement
Diameter - (mm) perpendicular to length
Height - (mm) with meat in shell
Whole weight - (grams) whole abalone
Shucked weight - (grams) weight of meat
Viscera weight - (grams) gut weight (after bleeding)
Shell weight - (grams) after being dried
Rings - number of rings

For classification, we begin by replacing the Sex values M, F, and I by 0, 1, and 2 respectively. Then replace data in rings column into four groups based on the ring age:

Class 1: 0 – 7 years
Class 2: 8 – 10 years
Class 3: 11 – 15 years
Class 4: >15 years

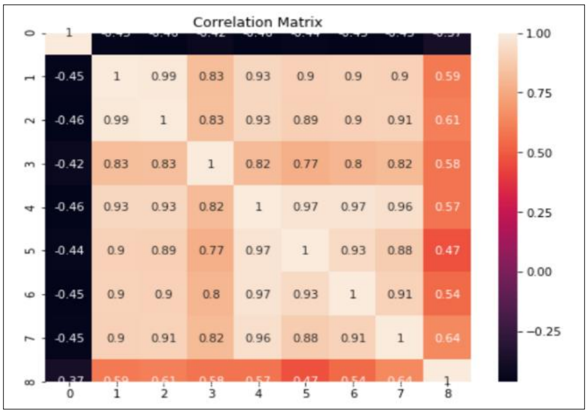
The result of processing is visualized by a bar chart. Nearly half of them are grouped into class 2: 8 - 10 years.

Figure 1



Obtain the correlation matrix by a heatmap plot from the seaborn library for processed dataset.

Figure 2



Split data using 60/40 percent train/test for this data set for modeling with optimizer.

B. Equations

F1-score will be used to determine the accuracy of the precision and sensitivity:

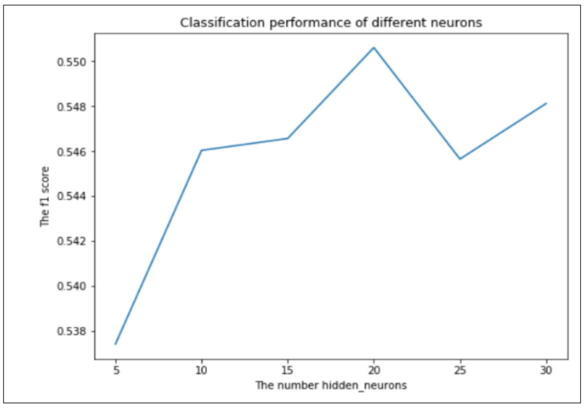
$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

III. Results

A. Effect of the number of hidden neurons

Experiment with different number of hidden neurons (5, 10, 15, 20, 25, 30) for a single hidden layer with max iteration 1000. Perform prediction in testing set to see the classification result. Then visualize the f1 score of each neuron. From the figure 3, the optimal number of hidden neurons should be chosen at 20.

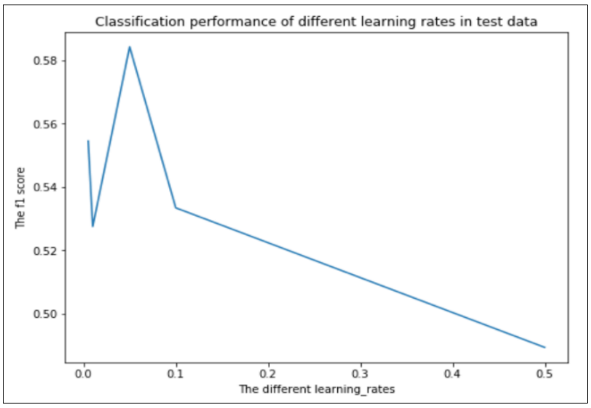
Figure 3



B. Effect of learning rate

The second experiment is testing with different learning rate (0.005, 0.01, 0.05, 0.1, 0.5) in case of SGD, using 20 hidden neurons. The line chart below shows that the f1 score of the classification performance is highest when rate equals to 0.05, at above 0.58.

Figure 4



C. Effect on a different number of hidden layers

Fit classifier with single and double hidden layers separately at layer sizes (20) and (20, 20). Store prediction results and measure the accuracy. The F1 scores of the single-layer neural network and the double-layer neural network are 0.54556 and 0.56467 respectively. The result shows that the later one is more precise when other parameters hold the same in test data.

D. Compare the effect of Adam and SGD

Train neural networks with the SGD and Adam optimizers on 2 hidden layers of 20 hidden neurons as selected before. The Adam optimizer fits the training set better at an accuracy of 0.57213, while the accuracy of SGD is 0.5329.

E. Best model using confusion matrix and ROC

Apply all the optimal choises that we have experimented:

Hidden neurons = 20;
Learning rate = 0.05;
Hidden layers = 2;
Solver = Adam;

The best model can be evaluated using a confusion matix for this classification problem. Table 5 reported the precision score, accuracy score, recall score and f1 score to show the accuracy clearly for the model.

The confusion matrix is :

257	77	3	0
64	496	146	1
7	184	326	10
1	11	76	12

Table 5

Precision	Accuracy	Recall	F1
0.6351	0.6529	0.5507	0.5611

The ROC curves for each class for a single experiment run are visualized below. For comparison, the axis is the same for each graph and the AUC data is shown in the title.

Figure 6.1 AUC = 0.72316

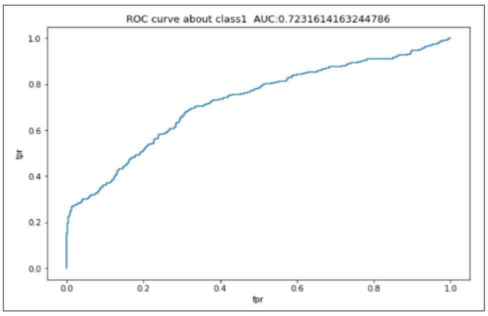


Figure 6.2 AUC = 0.47896

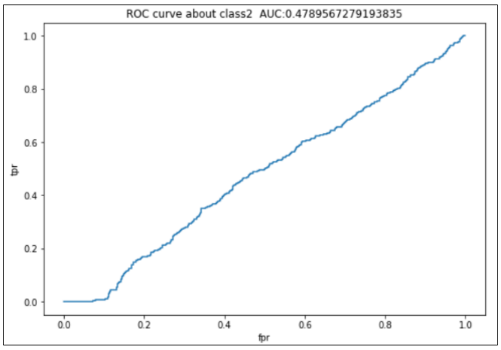


Figure 6.3 AUC = 0.38284

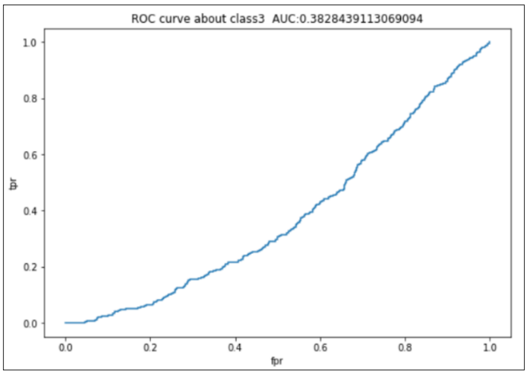
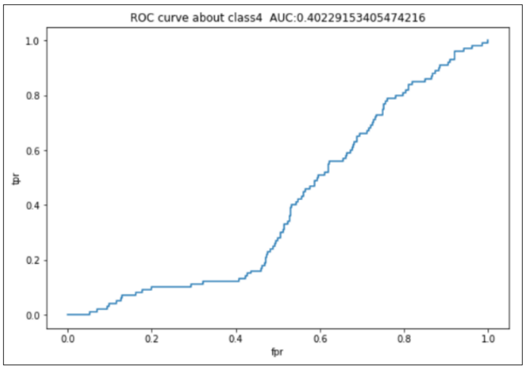


Figure 6.4 AUC = 0.40229



IV. DISCUSSION

From the measurement for the classification problem using experimentally optimal model, we observe that the model is capable to distinguish between classes. Note that the accuracy and AUC are both not as high as expected, because there are some limitations of the optimizer over the dataset. In this paper, we chose Adam algorithm for the final model. However, it may have a weight decay problem and the dataset is not large enough compared to the huge number of abalone in total.

V. CONCLUSION

To sum up, we classified the dataset into four groups effectively by neural networks. In this experiment, the Adam was proved to be the better optimizer. The prediction accuracy of the optimal model is around 65% and the AUC of each classifier is up to 0.72. The test results were visualized in provided figures and tables. Further research can be done with respect to the dataset, such as regularization and adding dropout work during training to improve generalization of the neural network.

REFERENCES

[1] Prasoon Goyal, “What are the pros and cons of stochastic gradient descent versus Adam as optimisation algorithms used in Keras machine learning?”, Quora. Aug 18, 2018.

[2] Vivek Kumar, “Why are neural network not the answer to everything?”, Analytics Insight. November 21, 2019.

[3] Diana Ramos, “Real-Life and Business Applications of Neural Networks.”, Smartsheet. Jul 19, 2021.

[4] Layan Alabdullatef. Complete Guide to Adam Optimization, Sep 2, 2020.

[5] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), pp. 359-366.

[6] Trenn, S. (2008). Multilayer perceptrons: Approximation order and a necessary number of hidden units. IEEE transactions on neural networks, 19(5), pp.836-844.