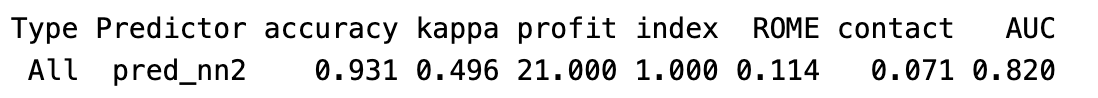
MKT 745 HW7 Zihao Li

1. In the data CASE\_WHERE2GO\_MODEL, use *buy* as dependent variable , all other variables except for *machine\_id* and *training* as exploratory variables to build a neural network classification model with decay=0.5 and size=1,2,3,4. Fill the table below with the AUC for both training and test data.

|  |  |  |
| --- | --- | --- |
| size | **Training data** | **Test data** |
| 1 | 0.933 | 0.666 |
| 2 | 0.998 | 0.669 |
| 3 | 1.000 | 0.668 |
| 4 | 0.999 | 0.660 |

1. What is the optimal size? Why? [5 points]

Optimal size will be 2 because it has highest AUC for test data and AUC for training data did not show overfit.

1. Build the model with decay=0.5 and size equals to the optimal size in b). Show the AUC and gain plot for All data. [5 points]

A graph with a line and a red line

Description automatically generated

1. Use the model in c) to predict the probability of purchase in data CASE\_WHERE2GO\_POOL. Show the average predicted probability for all candidate users. [10 points]

A screenshot of a phone

Description automatically generated

1. Show the distribution of the predicted probability for all candidate users. Do you observe any pattern when comparing the density of users who have high purchase probability with those who have low purchase probability? [10 points]

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

User with high connection speed and is not in the country of origin have higher predicted probability.

1. Please find evidence to support or oppose the claim below:

Users who have children have higher chance of booking hotel online

[hint: show model free evidence such as figure or statistics and then use t-test to support the model free evidence] [10 points]

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

The p-value is lower than 0.05 which means null hypothesis Is being reject. And people who have children have lower number of transactions of online hotel booking.

Now you have two alternative cutoffs to determine the list of final users: 0.3 and 0.7

1. If this campaign is charged by cost per user and it is expensive, which cutoff would you select? Why?

I would select 0.7. Because it will select lower amount of user and focus more on people who have higher purchase probability which can save money and make the result more meaningful.

1. How many final users and what is the average predicted probability for the final users in the condition g)?

A screenshot of a computer

Description automatically generated

1. If the users obtained from this campaign are the final customer base for a new and important service, which cutoff would you select? Why? [10 points]

I would select 0.3. Because it will include more users and will help to find out pattern. And also it will help to identify which kind of customer will be interest and help to understand in the future which kind of customer we should focus on to higher have higher ROI.

1. How many final users and what is the average predicted probability for the final users in the condition i)? [15 points]A screenshot of a social media account

   Description automatically generated