FINAL PROJECT Q&A

Instructions

The President of the bank asks you to design and implement a predictive model to determine which future applicants should be approved for a credit card and which rejected.

Your new goal is to develop a binary classification process for “approve” or “reject” designed to *maximize total bank profits*. (Put another way, you want to maximize average profits per applicant, including in the total number of applicants those that are rejected).

**Step-By-Step Assignment Instruction**

**Assignment Details:**

* Assume you have available to you the six standardized data inputs only (no Exotropia scores). Design your model on the Training Set only.
* You may adapt your own previous binary classification model for default, the example binary classification model for default given in the Learning Points for quiz 1 and 2,
* You may also use the linear regression model for profitability, or create a new model for this purpose.
* No methods outside the scope of what has already been taught in this course need to be, or should be, used.

**Note an important change in the bank’s assumptions:**

* The bank has learned that not all defaulters are unprofitable and not all non-defaulters are profitable. Therefore, you should no longer use the bank’s “old” estimates for rate of default (25%), average losses per defaulter (-$5,000) average profits per non-defaulter ($2,500), and average profits per applicant when using no model ($625).
* Instead, the *new* relevant *profitability metrics* assumptions the bank gives you to use are:
* The proportion of applicants that are unprofitable (again 25%, but a different 25%), average losses per unprofitable customer (-$4,900), average profits per profitable customer ($4,000), and average profits per applicant when using no model ($1,775).
* A “perfect” model that excluded all unprofitable and included all profitable customers would have an estimated profitability per applicant of ($4,000) \*.75 = $3,000

## ANSWER AND EXPLAINATION

**1. What is your predictive model?**

a. Describe the arithmetic clearly so that another learner could implement your model on new standardized input data if they wished.

b. Give an example of the score you would assign the following applicant, whether they would be approved or rejected for a credit card and why.

*Step1: Use Linest Function  
Step2: Use the formula which is: -0.07\*Automobile Debt + (-0.19)\*Credit Debt + (-0.08)\*Income + 0.03\*Years at address + (-0.19)\* Years At Employ + (-0.02)\*Age + 0.25  
So my predictive model has the formula as above and I would call this Score since it is used on the question below.  
I got AUC for training set is 0.84*

**2. Give an example of the score you would assign the following applicant, whether they would be approved or rejected for a credit card and why.**

*Step1: Base on TP, FN calculate loss from defaulters: -4900\*FN (1)  
Step2: Base on TP,FN calculate profit from non-defaulters: 4000\*TN (2)  
Step3: (1) + (2) = Total Profit  
Step4: Total profit / 200 = Profit per applicant >>> Max Profit per application  
Step5: Calculate max profit per threshold = IF(profit per app = max profit per application, Score,"")  
>>>>> Threshold for give card is 0.34  
So anyone above 0.34 is accepted and below would be rejected*

**3. What would the bank’s *average profit per applicant* be (net profits divided by 200) when using your predictive model on the Training Set?**

*I calculate total profit= (1)+(2) (see above) = 431400  
Average profit per application = 431400/200 = 2157*

**4. What is the *incremental financial value* per applicant of your model over no model on the Training Set?**

*I take the formula as: My model profit per app - No model profit per app = 2157 - 1775 = 382*

5. **Evaluate your model on the Test Set data. How confident are you that your model does not over-fit the Training Set data?**

A. Choose between three broad degrees of confidence: “very” “somewhat” or “not at all.” (Note that “not at all” is still an acceptable answer if you give persuasive reasons for why you chose this answer).

B. Explain the evidence your degree of confidence is based upon. Your explanation should include the test set profits and training set profits per applicant.

How much confidence to have in the model must relate to the relationship between the profits-per-applicant on the Training Set and the Test Set

*Very confident*

*Training Set: AUC: 0.84*

*- Profit: 407000*

*- Profit per app: 2157*

*Test Set:*

*-AUC: 0.86*

*-Profit: 435300*

*-Profit per app: 2176.5*

*>>> the difference between Training Set and Profit Set is low (around 5.6%)*