CS 584: Machine Learning

Pradyot Mayank(A20405826)

Spring 2019 Assignment 3

# Question 1 (50 points)

You will use the CART algorithm to build profiles of credit card holders. The data is the CustomerSurveyData.csv. The analysis specifications are:

**Target Variable**

* **CarOwnership**. The type of car ownership. This variable has three non-missing categories which are *Leased*, *None*, and *Own*.
* Drop all missing values in the target variable.

**Nominal Predictors**

* **CreditCard**. The type of credit card held. This variable has five categories which are *American Express*, *Discover*, *MasterCard*, *Others*, and *Visa*.
* **JobCategory**. The category of the job held. This variable has six non-missing categories which are *Agriculture*, *Crafts*, *Labor*, *Professional*, *Sales*, and *Service*.
* Recode all the missing values into the *Missing* category.

You will use the Entropy metric as the splitting criterion. You may want to write a Python program to assist you in answering the questions.

1. (5 points). What is the Entropy metric for the root node?

**Root node Entropy=1.0744900777690842**

1. (5 points). How many possible binary-splits that you can generate from the CreditCard predictor?

**K=5**

**No of binary splits=2k-1-1=15**

1. (10 points). Calculate the Entropy metric for each possibly binary split that you can generate from the CreditCard predictor. List your answers in a table. The table should have three columns: an index of the split, the contents of the two branches, and the split entropy metric.



1. (5 points). What is the optimal split for the CreditCard predictor?

**Optimal Split entropy for card: 1.0708382285522746**

**['Others', 'Visa'] and ['American Express', 'Discover', 'MasterCard']**

1. (5 points). How many possible binary-splits that you can generate from the JobCategory predictor?

**K=7**

**No of binary splits=2k-1-1=63**

1. (10 points). Calculate the Entropy metric for each possibly binary split that you can generate from the JobCategory predictor. List your answers in a table. The table should have three columns: an index of the split, the contents of the two branches, and the split Entropy metric.



1. (5 points). What is the optimal split for the JobCategory predictor?

**Optimal Split entropy for Job: 1.0720111150297396 since it has the lowest entropy.**

**['Agriculture', 'Sales'] and ['Crafts', 'Labor', 'Missing', 'Professional', 'Service']**

1. (5 points). Between the CreditCard and the JobCategory predictors, which predictor will you choose for producing the second layer (i.e., depth 1) of your decision tree?

**Split Entropy**=**1.0708382285522746**

**We will choose the predictor Creditcard or producing the second layer of the tree** **['Others', 'Visa'] and ['American Express', 'Discover', 'MasterCard'] as the entropy is lowest from both the predictors.**

# Question 2 (50 points)

In 2014, Allstate provided the data on Kaggle.com for the Allstate Purchase Prediction Challenge which is open. The data contain transaction history for customers that ended up purchasing a policy. For each Customer ID, you are given their quote history and the coverage options they purchased.

The data is available on the Blackboard as Purchase\_Likelihood.csv. It contains 665,249 observations on 97,009 unique Customer ID. We are going to use the MNLogit function to build a multinomial logistic model to predict purchase likelihood of coverage A using three predictors. The target variable is **A** which have these categories 0, 1, and 2. The nominal predictors are (categories are inside the parentheses):

1. **group\_size**. How many people will be covered under the policy (1, 2, 3 or 4)?
2. **homeowner**. Whether the customer owns a home or not (0=no, 1=yes)
3. **married\_couple**. Does the customer group contain a married couple (0=no, 1=yes)

Please build a multinomial logistic model using and answer the following questions.

1. (2 points) Suppose you start with a model with only the Intercept term (i.e., without any predictors). How many parameters are in this model?

**For intercept only model we only have one parameter is involved**

1. (3 points) What are the marginal counts of the categories of the target variable A?

|  |  |  |  |
| --- | --- | --- | --- |
| **ToatlCount\_0** | **ToatlCount\_1** | **ToatlCount\_2** | **ToatlCount** |
| 143691 | 426067 | 95491 | 665249 |

1. (5 points) Without calling the MNLogit function, what are the maximum likelihood estimates of the predicted probabilities of this Intercept-only model? Show all the necessary steps and the estimates for the . (Hint: equate the first derivatives of the log-likelihood function to zeros for this Intercept-only model)

|  |  |  |  |
| --- | --- | --- | --- |
| **Total Observation** | **MLE** | **Odds** | **MLE Intercept** |
| 143691.0 | 0.215996 | 1.000000 | 0.000000 |
| 426067.0 | 0.640462 | 2.965161 | 1.086931 |
| 95491.0 | 0.143542 | 0.664558 | -0.408633 |

1. (3 points) What is the log-likelihood value of this Intercept-only model? (Hint: the log-likelihood function is )

**(tototalCount\_0 )loge(prob\_0) + (tototalCount\_1)loge(prob\_1) + (tototalCount\_2)loge(prob\_2)**

**143691 loge(0.215996) + 426067 loge(0.640462) + 95491 loge(0.143542)**

**Log Likelihood: -595406.7618844224**

1. (5 points) Next, you are asked to mathematically calculate the maximum likelihood estimates of the Intercept terms The convention is to set the Intercept term to zero for the target category A = 0, i.e., . (Hint: use the mathematical formula of the logit of (i.e., for this Intercept only model, then solve for the betas)?

**J = 1, 2, 3**

**β = [β10, β20, β30]**

**𝜋𝑖1 = Probability of 0= 0.21599581510081187**

**𝜋𝑖2 = Probability of 1= 0.64046244338586**

**𝜋𝑖3 = Probability of 2= 0.1435417415133281**

**βj =loge (𝜋𝑖𝑗/𝜋𝑖𝐽)**

**β 1 = loge (𝜋𝑖0/𝜋𝑖0) = loge(1) = 0**

**β 2 = loge (𝜋𝑖1/𝜋𝑖0) = loge(0.6404624433858/0.2159958151008) = 1.08693**

**β 3 = loge (𝜋𝑖3/𝜋𝑖1) = loge(0. 143541741513/0. 2159958151008) = -0.40863**

1. (5 points) Create and display a contingency table where group\_size, homeowner, and married\_couple are on the row dimension, and A is on the column dimension. The cell contents are the row percentages of the categories of A per each level combination of group\_size, homeowner, and married\_couple.



1. (2 points) Based on the contingency table in e), do you expect the separation or the quasi-separation phenomenon to occur when we build the multinomial logistic model which has group\_size, homeowner, and married\_couple as the predictors.

**Based on the contingency table in (f), we expect the Quasi-complete separation because none of the predictor completely define particular category of target variable.**

1. (5 points) Now, you will use the MNLogit function to build the multinomial logistic model which has group\_size, homeowner, and married\_couple as the predictors. What value of the target variable A is used by the MNLogit function as the reference category? Next, what is the log-likelihood value of this model? Finally, how many parameters (including the redundant ones) are in the model?  
   **Reference Category Value of A=0**

**Log-Likelihood=-591936.7938327907**

**Since there are 9 parameters and each parameter has its occurrence and nonoccurrence i.e. So,Parameters Involved=9\*2=18**

1. (10 points) What are the values of group\_size, homeowner, and married\_couple such that the odd Prob(A=1)/Prob(A = 0) will attain its maximum? What is the maximum odd Prob(A = 1)/Prob(A = 0) value?

**Predicted probability of 0: 0.1606721019540383**

**Predicted probability of 1: 0.7020727585496748**

**Maximum odd Probability: 4.36959963809093**

1. (5 points) According to the multinomial logistic model, what is the odds ratio for group\_size = 3 versus group\_size = 1, and A = 2 versus A = 0? Mathematically, the odds ratio is (Prob(A=2)/Prob(A=0) | group\_size = 3) / ((Prob(A=2)/Prob(A=0) | group\_size = 1).

**group\_size\_3A2=-0.107280**

**group\_size\_1A0=0.411100**

**group\_size\_3A2-group\_size\_1A0**

**Required Odds Ratio: 0.5954844518092098**

1. (5 points) According to the multinomial logistic model, what is the odds ratio for group\_size = 1 versus group\_size = 3, and A = 2 versus A = 1? Mathematically, the odds ratio is (Prob(A=2)/Prob(A=1) | group\_size = 1) / ((Prob(A=2)/Prob(A=1) | group\_size = 3).

**group\_size\_3A2=-0.026011**

**group\_size\_1A1=0.108286**

**group\_size\_1A1-group\_size\_3A2**

**Required Odds Ratio: 1.1437324577458428**