CS 584-04: Machine Learning

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?
2. (5 points). How many possible binary-splits that you can generate from the CreditCard predictor?
3. (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.
4. (5 points). What is the optimal split for the CreditCard predictor?
5. (5 points). How many possible binary-splits that you can generate from the JobCategory predictor?
6. (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.
7. (5 points). What is the optimal split for the JobCategory predictor?
8. (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?

# 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?
2. (3 points) What are the marginal counts of the categories of the target variable A?
3. (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)
4. (3 points) What is the log-likelihood value of this Intercept-only model? (Hint: the log-likelihood function is )

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)?
2. (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.
3. (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.

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?
2. (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?
3. (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).
4. (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).