MSIS 510
Individual Assignment 4

Individual Assignment 4
Due at the beginning of the class on Nov. 30
Total: 8 points

East-West Airlines has entered into a partnership with the wireless phone company Telcon to sell the latter's service via direct mail. "phone_sale.csv" is a sample of data, provided so that the analyst can develop a model to classify East-West customers as to whether they purchase a wireless phone service contract.

In this problem, we use **Logistic Regression** to classify the **target** variable (**Phone_sale**). The variable description is as follows:

Field Name	Description
Topflight	Indicates whether flyer has attained elite "Topflight" status, 1 = yes, 0 = no
Balance	Number of miles eligible for award travel
Qual_miles	Number of miles counted as qualifying for Topflight status
Bonus_miles	Number of miles earned from non-flight bonus transactions in the past 12 months
Bonus_trans	Number of non-flight bonus transactions in the past 12 months
Flight_miles_12mo	Number of flight miles in the past 12 months
Flight_trans_12	Number of flight transactions in the past 12 months
Club_member	Member of the airline's club (paid membership), 1=yes, 0=no
Any_cc_miles_12mo	Variable indicating whether member added miles on any credit card type within the past 12 months (1='Y', 0='N')
Phone_sale	Target variable indicating whether member purchased Telcom service as a result of the direct mail campaign

MSIS 510
Individual Assignment 4

1. Preprocess the data. Re-code the target value "**Phone_sale = Yes**" to be class 1. Transform **Any_cc_miles_12mo** to a categorical variable with levels renamed as 'Y' and 'N', and set the base level to 'Y'.

2. We choose to use only **Bonus_trans** and **Any_cc_miles_12mo** as predictors and drop other variables.

3. Partition the data using the pre-generated training indexes provided in "a4_TrainIndex.rda" (with seed 111 and 60% training vs. 40% validation). Run logistic regression on training data.
4. Find the fitted coefficient values. Write down the model:

   logit =

5. Use the above logistic regression model to calculate the probability for (**Phone_sale = Yes**) if:

   Bonus_trans = 50, **Any_cc_miles_12mo** = 'Y'

   Prob(Phone_sale = Yes) = _____

If the cutoff value is set to 0.3, is the above customer predicted to purchase a wireless phone service contract?

MSIS 510
Individual Assignment 4

6. Set the cutoff value to be 0.5. Write down the confusion matrix of the **validation** data, with 1 as the important class (using **positive = "1"**).

7. In our training data, what is the percentage of customers that purchased wireless phones? Our goal is to identify potential customers that will purchase a wireless phone as a result of the direct mail campaign. Based on the confusion matrix you got in step 6, is our current logistic regression model (with cut-off value 0.5) useful for this task? Why?

8. To deal with the above issue, should the cutoff value be increased or decreased? Why?