**Mini Project 2**

The data for this project comes from the mobile advertising space. In order to encourage consumers to install its app (e.g. a game), an app developer advertises its app on other apps (e.g., other games) through a mobile advertising platform. Consumers viewing these ads on these other apps can click on the ad to install the app from the developer. We will refer to the advertising app developer as the advertiser. See figure below.

Advertiser

Advertising Platform

.

.

.

Publisher 1

Publisher 2

Publisher k-1

Publisher k

Consumer

Consumer

Consumer

Consumer

.

.

.

Install

Not Install

The dataset for this project contains data about ads from one particular advertiser through multiple publishers. Each observation corresponds to one ad shown to a consumer on a particular publisher app. The observation contains information about the publisher id, consumer’s device characteristics, and whether the advertiser’s app was installed or not. The description of the variables are given below.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Type** | **Description** |
| publisher\_id | Num | Publisher Id |
| device\_make | Char | Device Manufacturer |
| device\_platform | Char | Phone OS Type (iPhone / Android) |
| device\_os | Char | Phone OS Version |
| device\_height | Num | Display Height (in pixels) |
| device\_width | Num | Display Width (in pixels) |
| resolution | Num | Display Resolution (pixels per inch) |
| device\_volume | Num | Device Volume when Ad was displayed |
| Wifi | Num | Whether WiFi was enabled when ad was displayed (Yes = 1, No = 0) |
| Install | Num | Whether Consumer Installed Advertiser’s App (Yes = 1, No = 0) |

The training dataset is “MobileAdsTraining.dta” and the test dataset is “MobileAdsTest.dta”.

1. The advertiser needs to determine how much to pay for placing ad depending on the publisher and on the consumer characteristics. The optimal payment is proportional to the probability that a consumer seeing the ad will install the ad. Construct a suitable model to estimate the probability. Report the AIC and SC for the estimated model in the training data (show the SAS output). Plot the ROC curve for this model for the test data, and report the area under the ROC curve.
2. The advertising platform would like to determine whether to show the ad from this advertiser depending on the publisher and consumer characteristics. Showing an ad to a consumer who would not install the app results in an inconvenience cost ‘C1’ to the consumer. Not showing an ad to a consumer who would have installed the app results in a missed opportunity cost ‘C2’ to the consumer. The platform would like to minimize the total expected cost for consumers. The ratio of C2 to C1 can be one of the following: 25, 50, 100, 200. For each of these cost ratios, develop a classification scheme to determine whether to display the ad or not. Explain the procedure that you would follow, and how you would determine the classification parameters based on the cost ratios.

Deliverables

* Project Report: For each question above, describe the model building and selection process that you followed, along with suitable tables and graphs as necessary.
* Project Presentation: Prepare a 10 minute presentation describing your approach and results.
* SAS code: Include a SAS file with detailed comments to reproduce all the results, tables and figures in the report and in the presentation. The code must be clearly labeled so that it is straightforward to see how to reproduce a particular result / table / figure. If the code will not execute for whatever reason, then points will be deducted. The code should assume that it will be executed in the folder containing the training and test dataset