

STT 811

In-Class Assignment 8

This problem will use the OJ dataset.

1. Create a target based on the Purchase field with numerical values of 0 and 1.
 - a. `OJ$Purchase1 <- OJ$Purchase`
 - b. `OJ$Purchase1[OJ$Purchase1 == "MM"] <- 1`
 - c. `OJ$Purchase1[OJ$Purchase1 == "CH"] <- 0`
2. Split the data into training and test datasets (with a 75/25 split).
 - a. `split_pct <- 0.75`
 - b. `n <- length(OJ$Purchase1)*split_pct # train size`
 - c. `row_samp <- sample(1:length(OJ$Purchase1), n, replace = FALSE)`
 - d. `train <- OJ[row_samp,]`
 - e. `test <- OJ[-row_samp,]`
 - f. `OJ_train_mod <- glm(data = train, Purchase1 ~ PriceDiff + LoyalCH, family = binomial)`
 - g. `test_pred <- predict(OJ_train_mod, test, type = "response")`
 - h. `train_cm <- confusionMatrix(as.factor(as.integer(2*OJ_train_mod$fitted.values)),
reference = as.factor(train$Purchase1))`
 - i. `test_cm <- confusionMatrix(as.factor(as.integer(2*test_pred)), reference =
as.factor(test$Purchase1))`
3. Build a logistic regression model for your target based on PriceDiff and LoyalCH.
 - a. How significant are the coefficients?
 - b. Compute the confusion matrix for both the train and test datasets. How do they compare?
 - c. Test the probability calibration for the training dataset.
4. Create bootstrapped 95% confidence intervals for the 2 coefficients (not Intercept)