Statistical Modeling Course

Logistic Regression Assignment

We will use the `Default` data set in the `ISLR` package for this assignment.

Problem 1

Fit a logistic regression model that uses `student`, `income` and `balance` to predict `default` . Interpret the coefficients.

Answer:

Problem 2

Using the validation set approach, estimate the test error of this model. To do this, perform the following steps:

- Write a function that takes 2 arguments: a formula and a dateset
- The function should do the following:
- Split the sample set intro a training set and a validation set.
- Fit a multiple logistic regression model using only the training observations.
- Obtain a prediction of default status for each individual in the validation set by computing the estimated probability of default for that individual, and classifying the individual to the `default` category if the estimated probability is greater than 0.5.
- Compute the validation set error, which is the fraction of the observations in the validation set that are misclassified.
- The function should return the validation error.

```
set.seed(2019)
```

Problem 3

Use your function from Problem 2 to repeat the process ten times, using ten different splits of the observations into a training set and a validation set, then get the average of these test errors. Comment on the results obtained.

```
set.seed(2020)
```

Answer:

Problem 4

Using your choice of goodness of fit test, determine if the logistic regression model is reliable for inference. Compare the model's descriptive and explanatory power from its predictive accuracy in Problem 3.

Answer:

Problem 5

Add your predicted probabilites to the Default data frame and call the column `preds`. Use the following code to plot the ROC curve. Calculate the confusion matrix and AUC. Comment on the results results.

```
library(plotROC)
Default <-
  Default %>%
  mutate(default_num = as.numeric(default)-1)

ggplot(Default, aes(d=default_num, m=preds)) +
  geom_roc(n.cuts = 6, labelround = 4) +
  geom_abline(intercept = 0, slope = 1)
```

Answer: