This homework runs logistic regression to predict the binary feature of whether or not a person was admitted to graduate school, based on a set of predictors: GRE score, TOEFL score, rating of undergrad university attended, SOP statement of purpose, LOR letter or recommendation, Undergrad GPA, Research experience (binary).

The data set was downloaded from Kaggle: https://www.kaggle.com/mohansacharya/graduate-admissions

The data is available in Piazza.

## Step 1 Load the data

\* Load the data

\* Examine the first few rows with head()

```{r}

# your code here

```

## Step 2 Data Wrangling

Perform the following steps:

\* Make Research a factor

\* Get rid of the Serial No column

\* Make a new column that is a binary factor based on if Chance.of.Admit > 0.5. Hint: See p. 40 in the book.

\* Output column names with names() function

\* Output a summary of the data

\* Is the data set unbalanced? Why or why not?

Your commentary here:

```{r}

# your code here

```

```{r}

# put the summary here

```

## Step 3 Data Visualization

\* Create a side-by-side graph with Admit on the x axis of both graphs, GRE score on the y axis of one graph and TOEFL score on the y axis of the other graph; save/restore the original graph parameters

\* Comment on the graphs and what they are telling you about whether GRE and TOEFL are good predictors

\* You will get a lot of warnings, you can suppress them with disabling warnings as shown below:

```

{r,warning=FALSE}

```

Your commentary here:

```{r,warning=FALSE}

# your code here

```

## Step 4 Divide train/test

\* Divide into 75/25 train/test, using seed 1234

```{r}

# your code here

```

## Step 5 Build a Model with all predictors

\* Build a model, predicting Admit from all predictors

\* Output a summary of the model

\* Did you get an error? Why? Hint: see p. 120 Warning

Your commentary here:

```{r}

# your code here

```

## Step 6 Build a Model with all predictors except Chance.of.Admit

\* Build another model, predicting Admit from all predictors \*except\* Chance.of.Admit

\* Output a summary of the model

\* Did you get an error? Why or why not?

```{r}

# your code here

```

## Step 7 Predict probabilities

\* Predict the probabilities using type="response"

\* Examine a few probabilities and the corresponding Chance.of.Admit values

\* Run cor() on the predicted probs and the Chance.of.Admit, and output the correlation

\* What do you conclude from this correlation?

Your commentary here:

```{r}

# your code here

```

## Step 8 Make binary predictions, print table and accuracy

\* Now make binary predictions

\* Output a table comparing the predictions and the binary Admit column

\* Calculate and output accuracy

\* Was the model able to generalize well to new data?

Your commentary here:

```{r}

# your code here

```

## Step 9 Output ROCR and AUC

\* Output a ROCR graph

\* Extract and output the AUC metric

```{r}

# your code here

```

## Step 10

\* Make two more graphs and comment on what you learned from each graph:

\* Admit on x axis, SOP on y axis

\* Research on x axis, SOP on y axis

Your commentary here:

```{r}

# plot 1

```

```{r}

# plot 2

```