Homework 3 (20 Points)

Due: 2022/10/22

Submission Guidlines

1. You need to write your homework in *Rmd/* ipynb/Rnotebook.
2. Please submit two files: one script file and one generated *pdf file.

In this assignment, you will explore Logistic regression with graduate school admission data, you are **not** allowed to use built-in logistic regression modules. Please implement the code by yourself.

1. Please import the dataset "admission.csv". The dataset has a binary outcome variable called 'admit' (1: admitted, 0: Not), and three predictor variables: 'gre', 'gpa' and 'rank'. The variable 'rank' takes on the values 1 through 4. This specifies the tier that the student's undergraduate insitution falls into. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. You can use the rank as numeric data.
2. Implement a logistic regression using gradient descent (with intercept) of admit as the outcome, and gre, gpa and rank as covariates. Report your results.
3. Implement a logistic regression using Iterated Reweighted Least Squares (with intercept) of admit as the outcome, and gre, gpa and rank as covariates. Report your results.
4. What is the coefficient of gpa in your regression? How do you interpret this coefficient? What is the coefficient of rank? How do you interpret this coefficient?
5. For the coefficient of gre, design a hypothesis test to see whether the coefficient is significant different from 0. (You can choose α-level as 5%)
6. Another option for designing the logistic regression model is to treat 'rank' as a categorical variable: you need to convert 'rank' into several new columns to represent different ranks. (You can think about perfoming one-hot encoding). Report your new model and discuss the two models you have derived.