Texas Christian University
COSC 40023 Spring 2025
Assignment 6

Due: April 14th, Monday, at 11:30 PM (Late submission NOT accepted)
Submission (two files, no compression): `assignment6.py` and `assignment6.R` through TCU Online

Download the dataset called "Social_Network_Ads.csv". This is the same dataset as what we used in class. Build and visualize logistic regression models in both Python and R. Two files named `assignment6.py` and `assignment6.R` should be created. Here are some additional requirements.

1. Add degree-2 polynomial features. The added features should contain degree-2 age, degree-2 salary, and age times salary.
2. `25%` of the data should go to the test set. In addition, `random_state` must be set to `0` in Python and `seed` must be set to `123` in R.
3. Feature scaling is required for both Python and R.
4. Train your model based on the training set. Then, print out the confusion matrix and accuracy based on the test set.
5. A training set plot and a test set plot must be generated in each programming language. The style of the plots should be identical to the one used in class. In each plot:
    - The horizontal axis should be scaled age and the vertical axis should be scaled salary.
    - For background, use light red to represent the predicted region of "not purchased" and use light green to represent the predicted region of "purchased".
    - Use red dots to represent "not purchased" observations and use green dots to represent "purchased" observations.
    - Have proper title and axis labels.
6. Have sufficient single-line and multi-line comments.