CSE353 - HW4

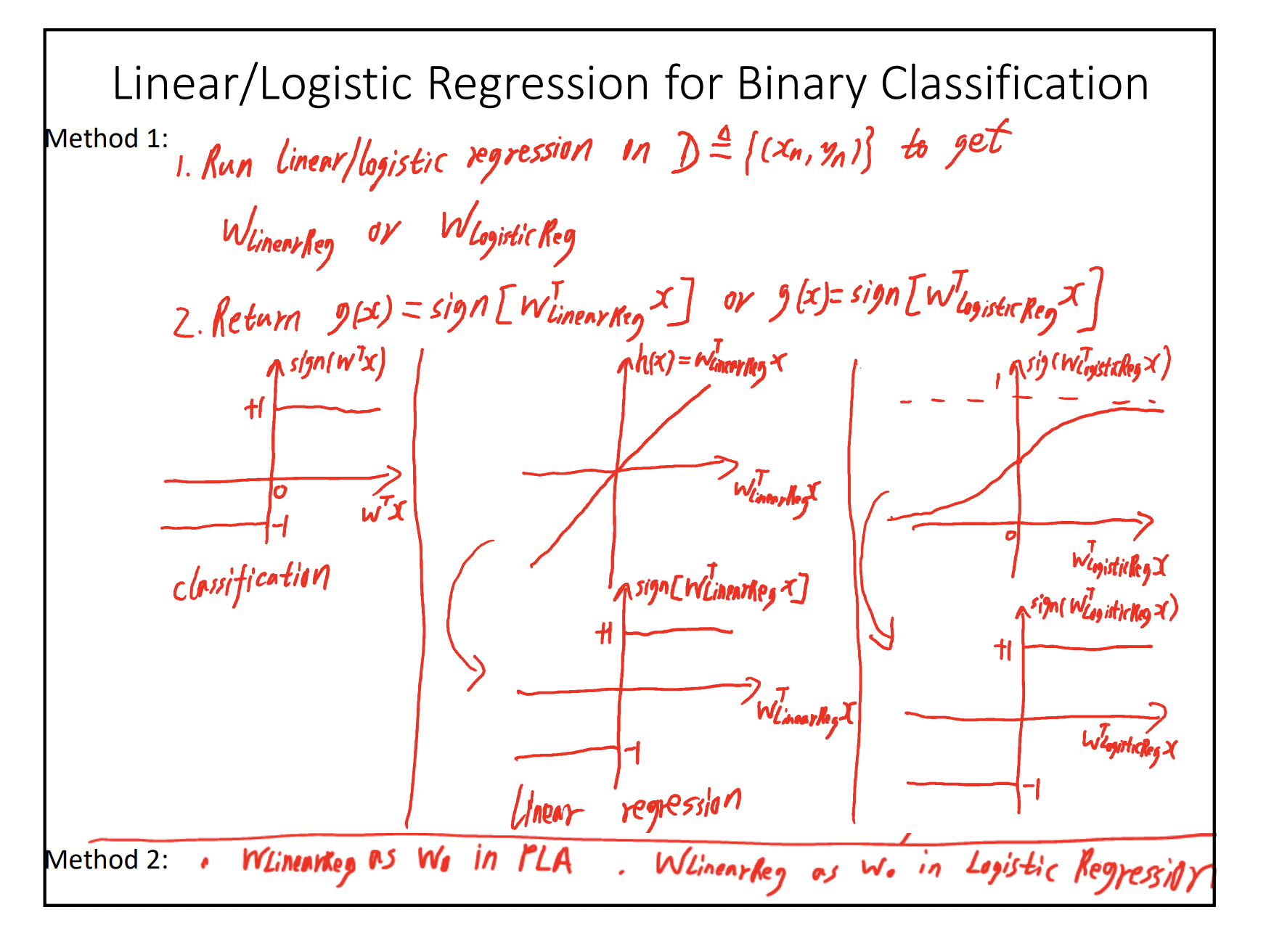
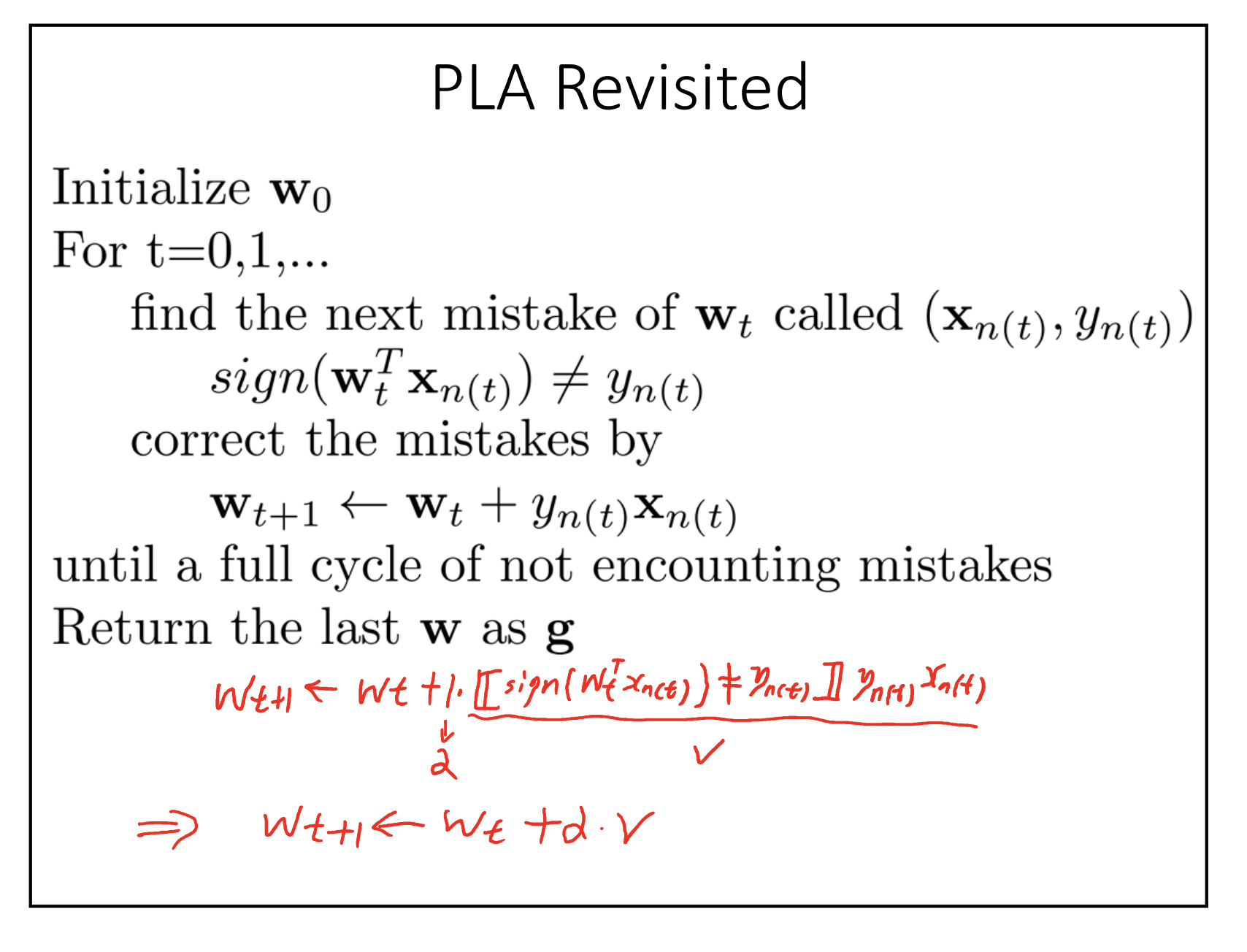
Yajie Wang

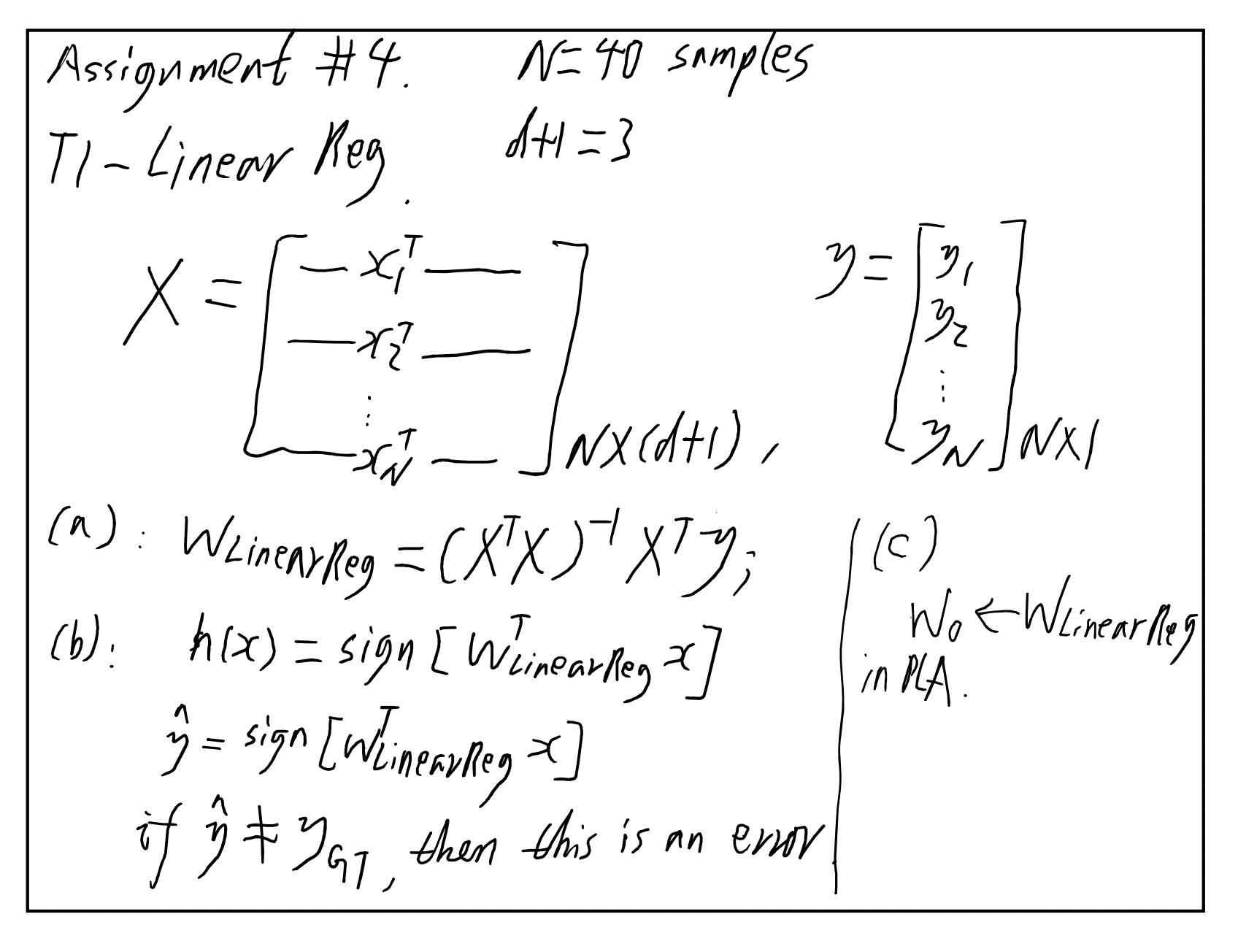
Introduction

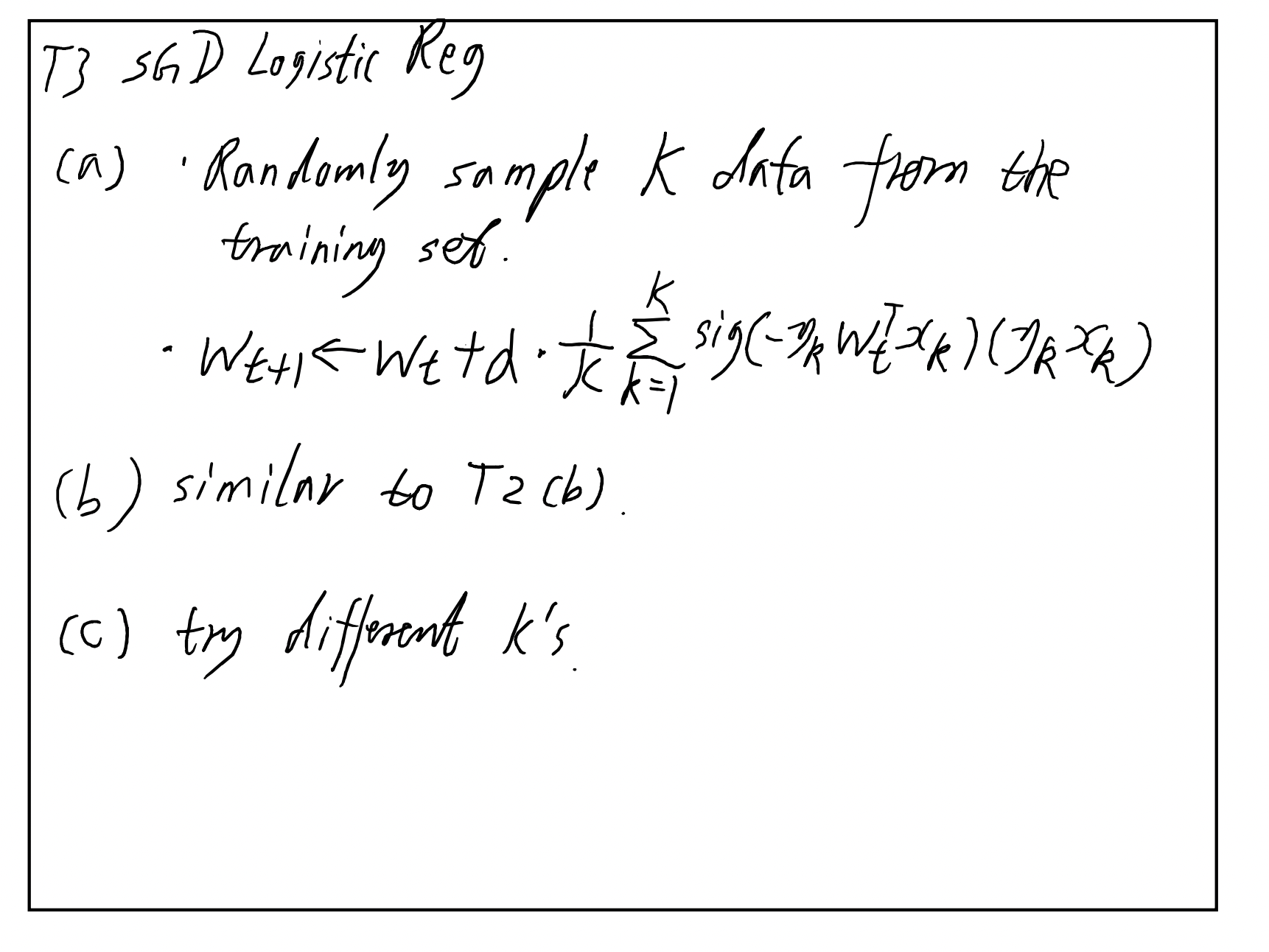
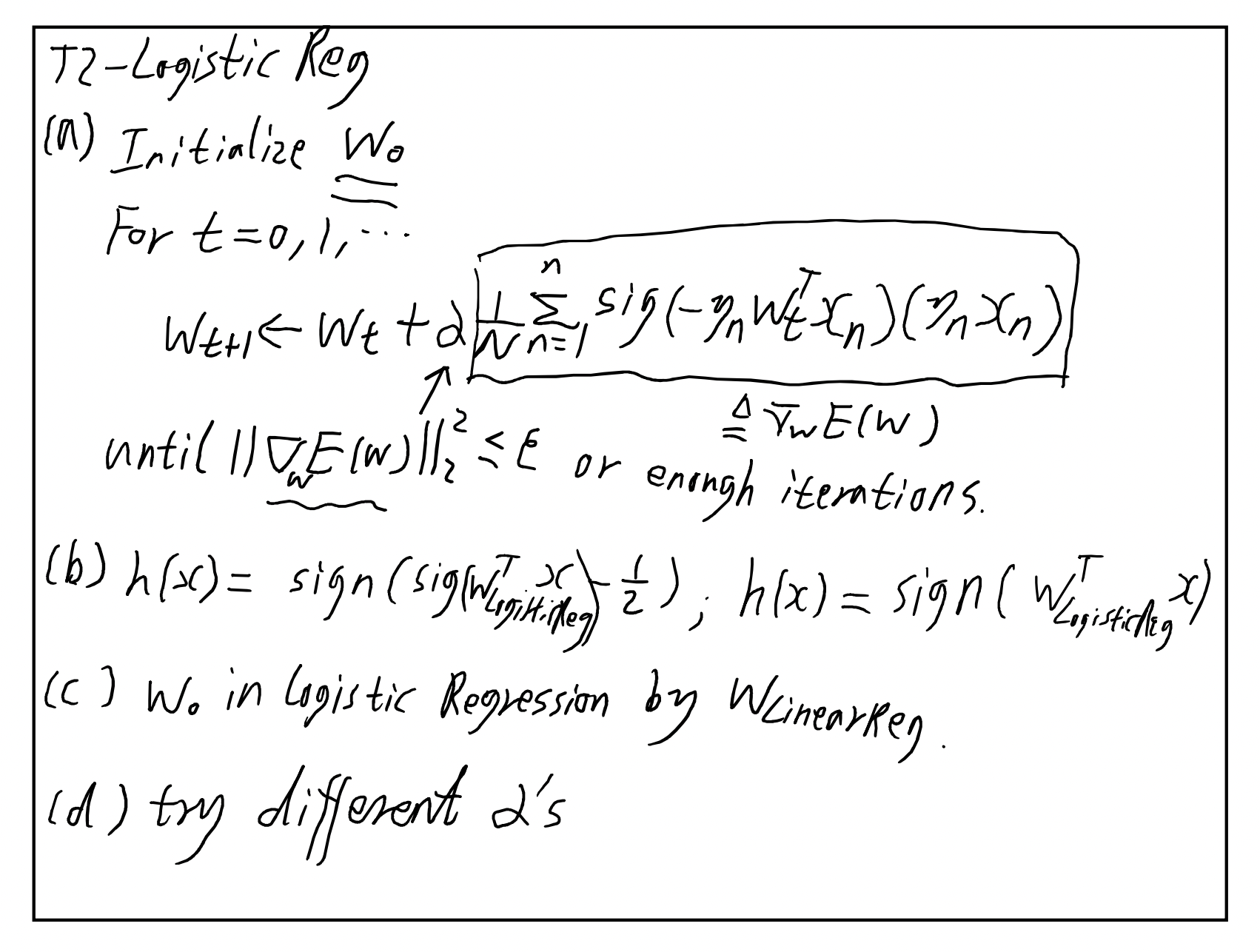
In this assignment, we implement the Linear Regression algorithm, Logistic Regression algorithm, Logistic Regression algorithm with Stochastic Gradient Descent (SGD) on the given dataset and obtain the corresponding w's. We apply these different w's to the dataset for binary classification and use them as the initializations to train the learning algorithms. We also plot the training data and the decision boundary learnt from the algorithm to visualize the results. By observing the results, we can see the pros and cons of these algorithms. For example, we compare the computational cost to see which algorithm is more efficient.

Method

The method I used in this assignment is based on what I learned in class. I implement the algorithms by following the instructions shown in the lecture slides. I define some get functions to calculate and return the values needed. I mainly use for loops and while loops to implement my algorithm.

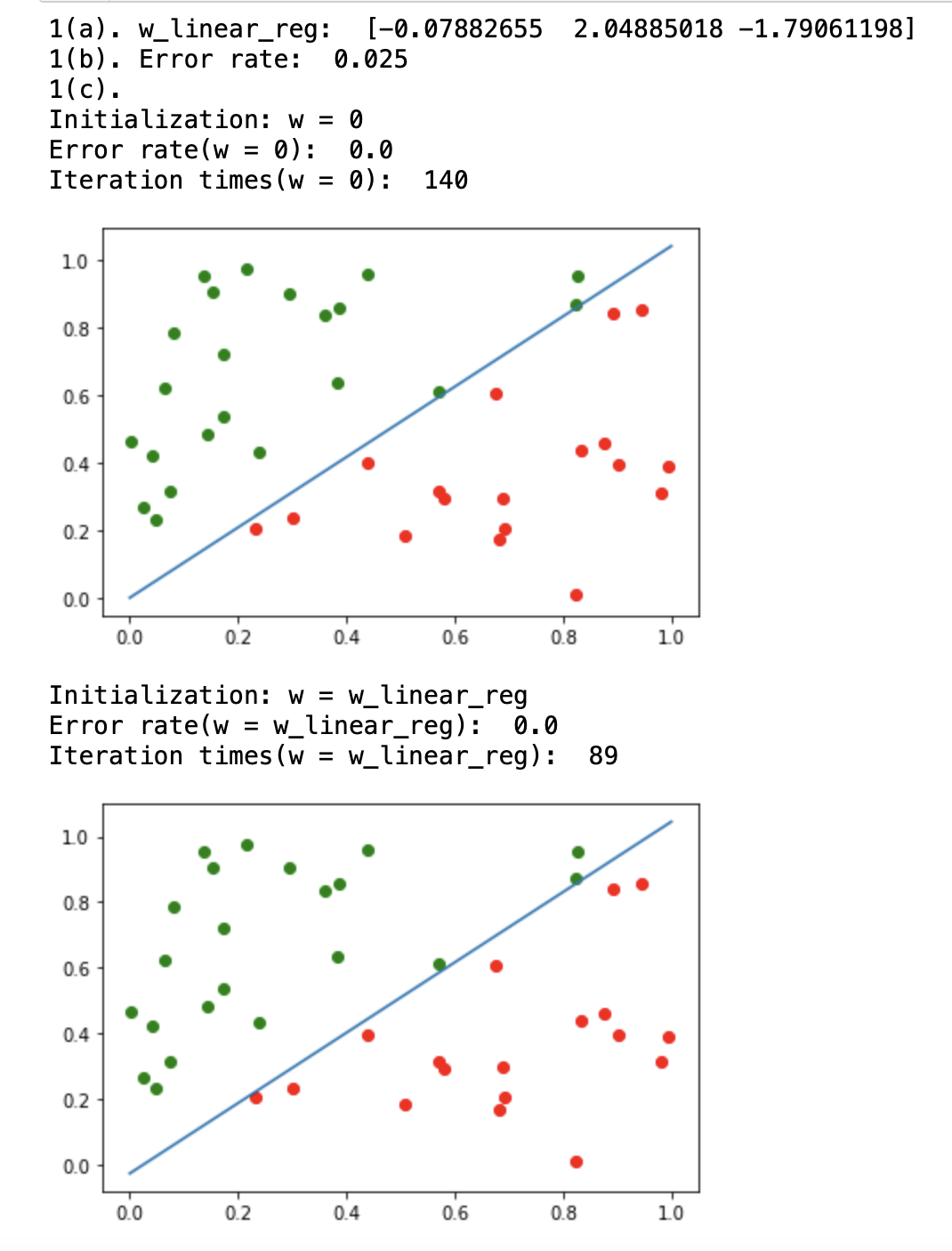




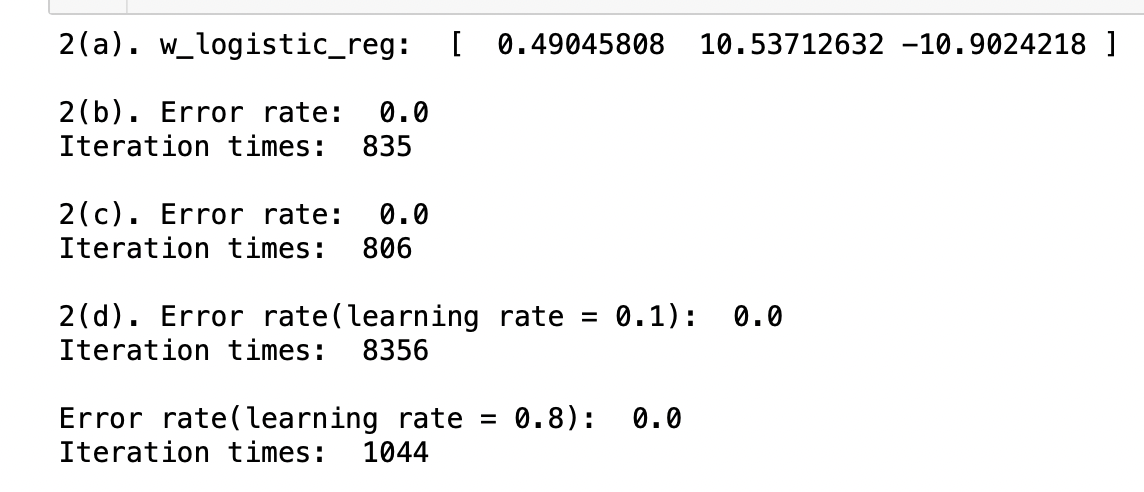


Experiments

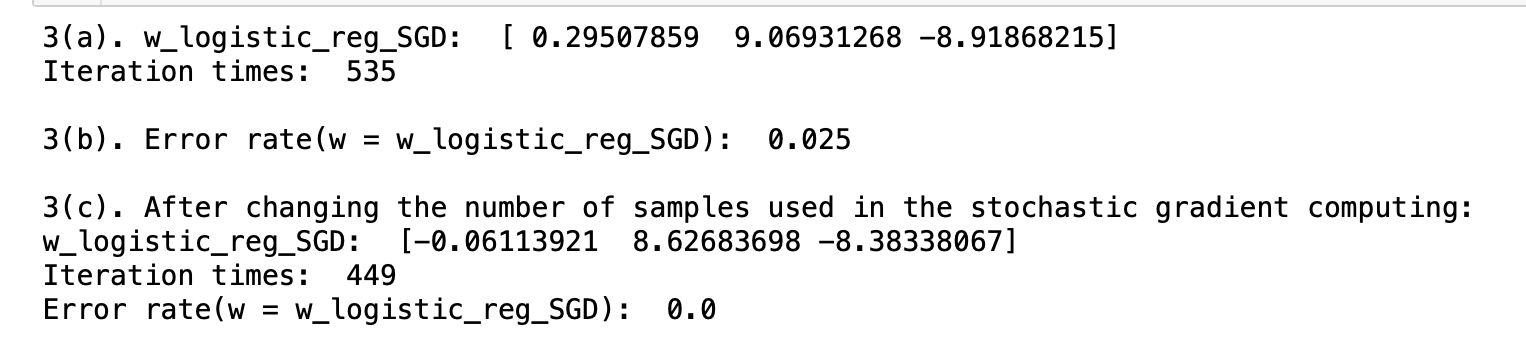
Part 1:



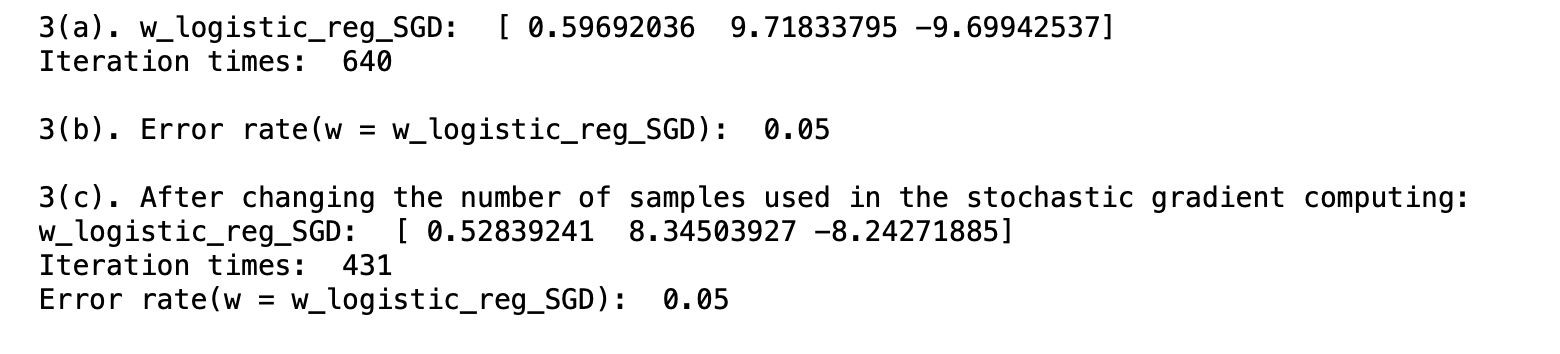
Part 2:



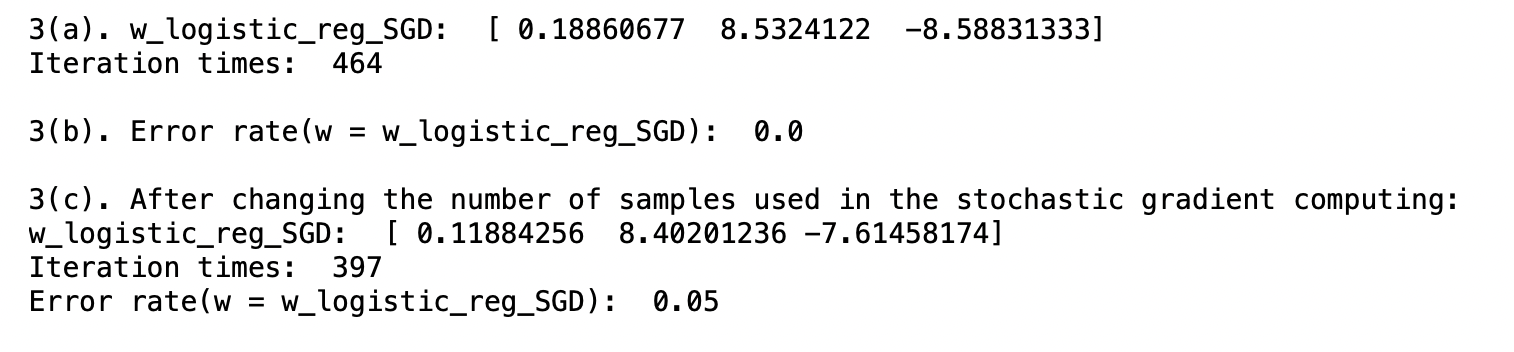
Part 3: three runs



run 1



run 2



run 3

Discussions

The result shows the Linear Regression algorithm, Logistic Regression algorithm, Logistic Regression algorithm with Stochastic Gradient Descent (SGD) all work as expected. All error rates are very low because the training data is small and linear separable. Although the training data is small, we can still see the difference in computational cost.

It’s more efficient to use 𝒘𝐿𝑖𝑛𝑒𝑎𝑟𝑅𝑒𝑔𝑟𝑒𝑠𝑠𝑖𝑜𝑛 as the initialization on 𝒘𝑃𝐿𝐴 to train our perceptron learning algorithm than other initializations like 𝒘𝑃𝐿𝐴 = 𝟎.

It’s more efficient to use using 𝒘𝐿𝑖𝑛𝑒𝑎𝑟𝑅𝑒𝑔𝑟𝑒𝑠𝑠𝑖𝑜𝑛 as the initialization in the logistic regression than an initialization with the zero vector. It is more efficient when the learning rate is larger. The SGD Logistic Regression algorithm is more efficient than the regular Logistic Regression algorithm. But the result may be less stable. And if we change the number of samples used in the stochastic gradient computing, it is more efficient when a smaller number of samples is used. And the result may be less stable.