Programming Assignment 2: Logistic Regression

Introduction

In this assignment you are required to use logistic regression, from the popular "scikit learn" python library, to implement a model that predicts some binary target (class) based on multiple features (at least 5 features). You will also write the code for the same problem from scratch, rather than using the ready to use logistic regression from scikit learn, to gain a better understanding on how things work under the hood. You are required to find such dataset for your assignment to use it in your code. A simple and clear description of the dataset, its source, and link to it should be included. Similar to previous assignments, you have to use it in CoLab through a link generated from your Google Drive.

Part 1: Using the sklearn library

Create a model and fine tune its hyper parameters, such as the solver, and C which is 1/regularization parameter, using grid search and k-fold cross-validation. You will need to report all metrics we have learned in class: confusion matrix, accuracy, precision, recall, ROC and AUC. Of course you need to leave a portion of the dataset only for testing. Use feature scaling for your data in this part using scaling method implemented in sklearn.

Part 2: Using the sklearn library

Create another Colabe notebook and write your own implementation of the logistic regression as we have learned it in class. Report same metrics for this model and compare them to the results of the sklearn model. Implement feature scaling for your data in this part from scratch.

In both files of 1 and 2, use proper documentation and make your code easy to get. I expect that each student will have different dataset and different details in coding and parameters. Cheating will not be tolerated.