



ECEN 4493 AI in Engineering Spring 2023



Homework Assignment #3

Statement of Academic Honesty:

For this homework, I make the following truthful statements:

- I have not received, I have not given, nor will I give or receive, any assistance to another student taking this quiz.
- I will not plagiarize someone else's work and turn it in as my own.
- I understand that acts of academic dishonesty may be penalized to the full extent allowed by the University Student Conduct Code, including receiving a failing grade (F!) for the course. I recognize that I am responsible for understanding the provisions of the University Student Conduct Code as they relate to this academic exercise.

Your Signature

Problem 1: Download the zip folder “Homework 3 Supporting Material.rar” from Canvas. It contains the MATLAB implementation of Logistic Regression. Open the “ex2.mlx” file and it will walk you through how logistic regression is coded. In Section 2.5, vary the regularization parameter γ and observe the changes to the decision boundary and the accuracy. Plot the decision boundaries for different γ and comment on why you observe such behavior.

Problem 2: In the downloaded zip folder, open the “Homework_3_Question_1.ipynb” file on Python environment of your choice (You need to upload it along with the datafile to google drive if you are using Google Colab). Complete the notebook by implementing the logistic regression functions. Upload the notebook to Canvas.

Problem 3: Now that you have coded the underlying functions of logistic regression, you can use the built-in libraries to implement it. Open the “Homework_3_Question_2.ipynb” file. You need to complete the notebook and perform the following tasks:

- 1) Describe the different parameters given to the *LogisticRegression* function. You can find the documentation on https://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.
- 2) Run the Logistic Regression for 5, 30 and 50 iterations and comment on the results and tabulate them. Provide reasoning on the behavior you have observed.
- 3) Upload the notebook.

Please save all submitted files (2 notebooks and 1 report) into a single standalone ZIP file and submit your ZIP file including all reference materials and citations, through **Canvas**

Assignments folder, “4493-homework3-Spring23 (assign 2-27-23, due 3-13-23)” by the deadline, March 13, 2023, 11:59pm.

For lambda equal to zero you get the best accuracy however the circle is not very uniform and tries to avoid the not accepted parameters. For lambda equal to 1 you get a relatively uniform circle that takes mostly the correct values but also has a few bad ones. For lambda equal to 10 it accepts almost everything regardless of whether it is good or bad and only has a 50% accuracy. The value of 100 accepts absolutely everything the boundary line can no longer be seen and the accuracy is a grand total of 7%. Lambda regulates the boundary and the higher the lambda value the less tightly bound the constraints leading to lower accuracy. Lambda of 1 gives the best overall circle however lambda 0 gives the best accuracy and data partitioning however it appears to be slightly overfitted, Therefore I would go with 1.

Problem 3:

The different parameters found in the sklearn logistic regression function are penalty, dual, tol, c, fit_intercept, intercept, scaling, class_weight, random_state, solver, max_iter, multi_class warm_start, verbose, n_jobs, l1_ratio,

Penalty: affects the penalty result and the type of penalty term.

Dual: type of formulation for L2 penalty and liblinear solver.

Tol: tolerance for the stopping criteria

C: affects regularization the smaller the value the stronger the regularization

fit_intercept: sets if a constant needs to be added in the decision-making function.

Intercept_scaling: used in liblinear to append instant vector.

Class_weight: assigns a weight value to determine the importance of a class to the learning process. Adjusts how focuses the different class labels will be.

random_state: how the data is shuffled

Solver: type of algorithm used to optimize the data set. Depending on the size some are better than others

Max_iter: is the number of iterations to converge.

Multi_class: controls loss minimizations and problem type

Verbose: affects output results

N_jobs: number of cores that are used in parallel

l1_ratio : affects the penalty parameters for elasticnet

The parameters for the sklearn logistic function affect the different ways the model processes its contents, data, and solver constants. Depending on the data you use you can mix and match settings to obtain more optimal results. You can also increase and decrease iterations to find more optimal solutions and fix the best fit.