

Project 3

Classification Using Logistic Regression

Due before midnight on October 18 (3^{n-1} points deducted per day late)

Problem Description and Data Set

For Project 3 we will revisit the problem from Project 1 of identifying whether a fish is TigerFish0 or TigerFish1. This time you are to use a logistic regression approach. You should use the same data file that was emailed to you for Project 1. The first line contains a single integer indicating how many sets of labelled data you have to work with. Each line after that contains three tab-separated entries. The first is a float representing the body length in centimeters, followed by a float representing the dorsal fin length in centimeters, then an integer identifying the fish as either TigerFish0 (with a 0) or TigerFish1 (with a 1).

Assignment

Using what you have learned in class and the notes, develop (from scratch in Python, no using a Logistic Regression library function!) a Hypothesis function that will predict type of fish given unseen data. You are free to use any model variation and any testing or training approach we have discussed.

What to Turn In Via Canvas

A pdf file (*lastname_firstname_P3.pdf*) that includes

- Problem Description
- Description of your data set along with a plot of the data.
- A description of your model and testing procedure, including
 - Initial values that you chose for your weights, alpha, and the initial value for J .
 - Final values for alpha, your weights, how many iterations your learning algorithm went through and your final value of J on your training set.
 - Include a plot of J (vertical axis) vs. number of iterations (horizontal axis).
 - If you did feature scaling, describe what you did.
 - Value of J on your test set.
- A confusion matrix showing your results on the test set.
- A description of your final results that includes accuracy, precision, recall and F1 values.
- A comparison of your results with a logistic regression approach as compared to your previous k-nearest neighbor approach (including kNN confusion matrix, accuracy, precision, recall and F1 values).

Note that we have covered many possible tools you could use such as plotting, model selection with training, validation and test sets, feature scaling, and/or regularization. However, a simple approach that works is better than banging your head against your desk because a more sophisticated approach is not working!

Also submit a *single python (lastname_firstname_P3.py) file* that repeatedly prompts for the body length and dorsal fin length (in centimeters) of a candidate fish and prints out the type to the screen. The program should terminate when you enter zero for both values.

Do NOT zip, tar or feather your files! Upload them to canvas with the proper names