

Term Project 2 – The “Fruits” Problem

Spring 2021

Number of students per group: (same groups as Project 1)

Due Date: Thursday 15/4/2021 at 11:59PM

The aim of this project is to implement the perceptron algorithm and attempt to build a classification system that targets multi-class problems. In doing so, you will get a hands-on experience on reducing a given multi-class dataset, solving any encountered data imbalance problems, and using the necessary multi-class evaluation measures. The results should be communicated in a 3-6 pages long (font size 12) detailed scientific report. For this project, **you may only use Python and the libraries numpy and matplotlib (and the included utils.py file without modifications)**. Please follow the steps below:

- 1) Included in your project .zip file is a dataset containing pictures of fruits. Each picture represents one of 13 fruits (see the folder names), and each image is a 100x100 jpg image. Your first task is to read the pictures as numpy arrays using the functions inside the utils.py file. Note there are two kinds ways to read the images: (1) as 1D numpy array ($100 \times 100 \times 3 = 30000$ features), or (2) as a 3D numpy array which has the dimensions $100 \times 100 \times 3$. At the end of the task you will need to be able to create a “dataset” that you can *train* and *test* on. You are free to create new meta-features from the 30000 existing ones as long as you do not use any external library than numpy. One example is to divide the image into 9 regions (3x3 equal squares) and then extract the mean, std, max, and min values for each color channel (red, green blue).

Below are example images of each fruit in the dataset: (can you guess what they are?)



- 2) Implement the perceptron algorithm:
 - a. A learning function that will return the optimal bias and weights. Note: this function should take number of epochs as a parameter.
 - b. A classification function that uses the output from 2(a) to classify instances
 - c. A classification function that takes an 100x100 image, and prints what the image has. Examples, “Apple”, “Lemon” ...etc.
- 3) Reduce the multi-class problem using the One-vs-All technique learned in class.
- 4) Using the dataset you reduced in (3), do the following:
 - a. Choose an appropriate split for training and testing.
 - b. Train a set of perceptrons as required by the One-vs-All technique, and indicate whether you used any oversampling, or subsampling techniques and write your justification.
 - c. Test the performance (Accuracy) of the system as you increase the number of epochs during induction.

- d. Experiment with different values of the learning rate η (you do not have to show analysis for this parameter).
- 5) Show the performance of your classification system using:
- a. Accuracy, Precision, Recall, and F1 (for each class separately).
 - b. Show the overall performance of your system (Overall Accuracy, Macro-Recall, Macro-Precision, and Macro-F1).
- 6) Write a 3-6 pages long report that documents your experiments and results. The following are expected to be mentioned in the report:
- a. The design of the features you created (if any).
 - b. The design of the learning function for the perceptron algorithm.
 - c. The classifier testing procedure being followed.
 - d. The performance while varying epochs (for at least 1 of the classes)
 - e. Provide the overall performance of your system (per class, and as a whole).
 - f. Show which two classes were the most confusing for your system.
 - g. Show example images that were incorrectly classified and indicate what your system thought they were.
 - h. Write a short paragraph about what you believe would happen if you used the All-vs-All approach.
- 7) Submit your report (.pdf) and your code (.zip) to blackboard as two separate files titled ML_Project2_Group_**X.pdf**, and ML_Project2_Group_**X.zip** (where **X** stands for your group letter).

I have reserved some of the pictures which are not included in your dataset. Towards the end of the semester, the top 3 performing groups will be announced 😊

Feel free to ask any questions by email, or during office hours.

Let the game begin!