Machine Learning MC886

University of Campinas (UNICAMP), Institute of Computing (IC)
Assignment #2, 2019s2, Prof. Sandra Avila

Objective

Explore classification techniques and come up with the best possible model to the problem, avoiding overfitting. In particular, build an object recognition system to accurately classify images using the CINIC-10 dataset (`https://github.com/BayesWatch/cinic-10`).

Activities (10 pts)

1. (1 pts) Perform Multinomial Logistic Regression (i.e., softmax regression). It is a generalization of Logistic Regression to the case where we want to handle multiple classes. **You should implement your solution**. What are the conclusions?

2. (3 pts) Move on to Neural Networks, using one or two hidden layers. **You should implement your solution**. Does the performance improve?

3. (0.75 pts) Implement two update rules used to optimize Neural Networks (for example, Momentum, Adagrad, Adam). For reference, see `http://ruder.io/optimizing-gradient-descent`. **You should implement your solution**. Does the performance improve?

4. (0.25 pts) Pick your best model and plot the confusion matrix in the **test set**. What are the conclusions?

5. (4 pts) Prepare a 4-page (max.) report with all your findings. It is UP TO YOU to convince the reader that you are proficient on Logistic Regression and Neural Network, and the choices it entails.

6. (1 pts) You should provide a single Jupyter notebook with your solution (in Python 3 code).

Dataset

CINIC-10 consists of 270,000 32×32 color images in 10 classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck). Don't worry, for this assignment we separated 100,000 images, with 80,000 images for training, 10,000 images for validation and 10,000 images for test.

Dataset Information:

- You should respect the following traininig/validation/test split: $80,000$ training examples, $10,000$ val examples, and $10,000$ test examples. Avoid overfitting.

- The data is available at: `https://tinyurl.com/y6fmvllh`.

Deadline

Saturday, **October 5th**, 11h59 pm.
Penalty policy for late submission: You are not encouraged to submit your assignment after due date. However, in case you did, your grade will be penalized as follows:

- October 6th 11h59 pm : grade * 0.75
- October 7th 11h59 pm : grade * 0.5
- October 8th 11h59 pm : grade * 0.25

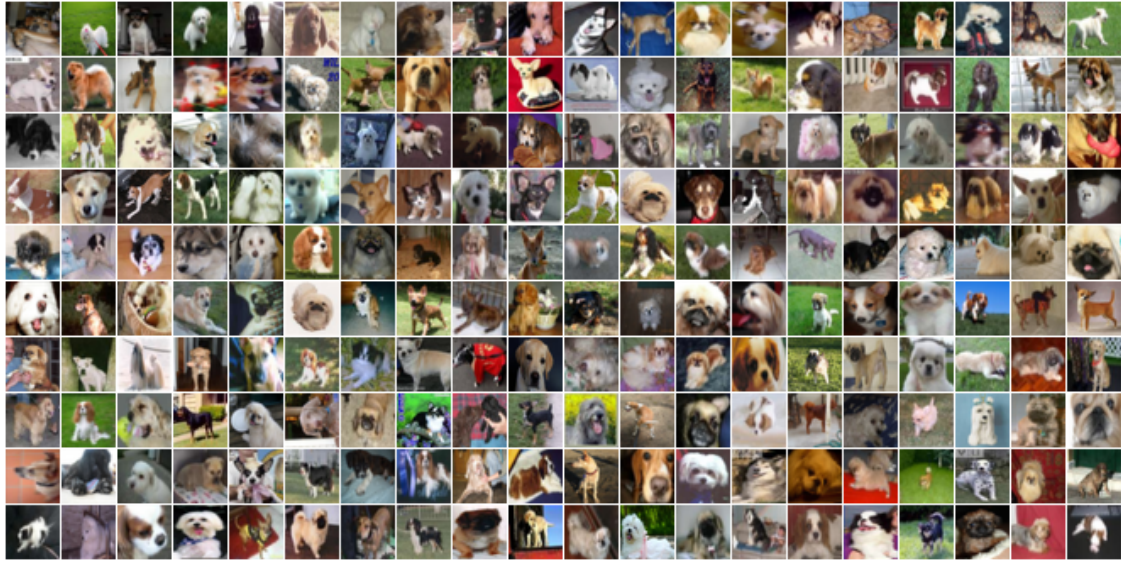Figure 1: The CINIC-10 dataset (dog class). Figure reproduced from `https://github.com/BayesWatch/cinic-10`.

Submission

On October 7th (Monday), bring your 4-page printed report. The template for report is available at `https://www.dropbox.com/s/nc6d89otr8ekvjd/report-model.zip`. Please, print on both sides of the page. The report should be written in Portuguese or English.

Submit a zip file, with the code and the report (PDF file), via Moodle.

This activity is NOT individual, it must be done in pairs (two-person group).