Machine Learning Mini-Project
May 20-24, 2019

Introduction

During this week you will be asked to work on the Kuzushiji-MNIST dataset.

Kuzushiji-MNIST is a drop-in replacement for the MNIST dataset (28x28 grayscale, 70,000 images, labelled into 10 classes), provided in the original MNIST format as well as a NumPy format.

This will be a contest between the different groups to produce a neural network that best classifies the images of the test set. The training set and test set respectively contain 60,000 and 10,000 images. You will have to create your validation sets out of the training set in order to optimize your hyper-parameters.

You are allowed to use any computational resources that are available to you but we recommend you use the facilities provided by Colab.

You will need to produce a neural network whose result on the test set can be reproduced. The program will also need to be properly documented. In addition you will need to produce a short report and a presentation.

Detail of Deliverables

We expect four deliverables from you:

1. **The code** that optimized your parameters and hyperparameters and created predictions on the test set. Your calculations will have to be reproducible, meaning that any user should be able to input the parameters: random seeds, hyperparameters, etc… that produce your optimal classification. The maximum code mark will be 50 points.
   25 points will cover "functionality", that is essentially the performance of the code on the test set. Based on the results of the different group, 15 points will be given to the teams who have achieved a minimum value of the test set accuracy, then the 10 remaining points will be based on the ranking of the test set accuracy obtained by the eight teams.
   The other 25 points will cover "sustainability", that is software documentation, reproducibility, state of submitted repository, presentation and accessibility of results e.g. README, organisation of notebooks / python scripts / plotting.
   The code should be in a repository with all the resources allowing reproduction of your final results.
2. **The short report** (maximum 2,500 words excluding figures) describing your neural network and the results you obtained on the test set:
   - Abstract (3 points)
   - Introduction (2 points)
   - Data pre-processing (3 points)

- Neural Network Architecture (5 points)
- Training Approach (3 points)
- Validation Approach and Hyperparameter Selection (7 points)
- Presentation and Discussion of results obtained on the Test Set (5 points)
- Conclusion (2 points)

The maximum report mark will be 30 points. It is recommended that you start writing your report early, and not wait until Friday morning to do so.

3. **A presentation** describing your results. This will be a 10' team presentation by one or two presenters from your team. During the following 5' Q&A session we will check that all team members are knowledgeable about the code and the results. The eight teams will attend all the presentations. The maximum presentation mark will be 10 points. The presentation is not marked for its technical content but its general quality as a presentation.
4. **Each team member** will have to fill the peer assessment form. This will be used together with our observations of the exchanges with team members during the week to give a maximum mark of 10 points. This last mark, contrary to the three others above, will be different for each team member.

Planning

Note that, in addition to room 1.51, room 3.03 will also be available for the groups between Monday afternoon and Friday morning. You are encouraged to also use room 3.03.

Monday
Morning at 9.00 am in Room 1.51: Presentation of the mini-project.
Afternoon from 2 pm – 4 pm: 15 minutes meeting of each team with Olivier or Lukas in room 2.34. Order of the groups:
2.00: Entropy
2.15: Convolution
2.30: Gradient
2.45: MaxPool
3.00: ReLU
3.15: Backprop
3.30: Sigmoid
3.45: Softmax
First two submissions to the leader board (submissions counter reset at 1 am, but you do not have to work until that time!).

Tuesday-Thursday
The agenda for these three days is the same:
10-11 am: possible interactions with Lukas in Room 1.51.
2-3 pm: possible interactions with Simon in Room 1.51.
Afternoon from 2 pm – 4 pm: 15 minutes meeting of each team with Olivier or Lukas in room 2.34.
Order of the groups for Tuesday:
2.00: Softmax
2.15: Sigmoid

2.30: Backprop
2.45: ReLU
3.00: MaxPool
3.15: Gradient
3.30: Convolution
3.45: Entropy
Order of the groups for Wednesday: same order as Monday
Order of the groups for Thursday: same as Tuesday.
Two Submissions to the leader board.
On Thursday, selection by each group of their two final submissions to the private leaderboard before 10 pm.

Friday
Report [.docx or.pdf] and presentation [.pptx or.pdf] to be submitted by noon.
In the afternoon, all groups attend all presentations from 2 to 4 pm in Room 1.51.
Format: 10' presentation + 5' Q&A by each group.
4-5 pm: general discussion and debriefing for all teams.