Assignment 4 - Multi Layer Perceptron

Advanced Topics in Neural Networks

24 October 2023

Homework

Develop a training pipeline with the following components:

1. A custom Dataset class that you must implement, that reads images from an image folder and applies transformations to them. (1 point)
2. Two PyTorch DataLoaders, one for training and one for validation. (1 point)
3. Training functions. (1 point)

   The training functions should have the following structure:

   - **run**, receives n, the number of epochs and calls the training and validation functions n times, while also updating the training and validation metrics.
   - **train**, trains the model for one complete iteration through the training DataLoader.
   - **val**, validates the model using the validation DataLoader.

   Other structures are also accepted, as long as they are readable, efficient and sensible.

Report training and validation metrics and create a graphic which illustrate their evolution (1 point).
Write an efficient implementation of the training pipeline (resort to caching the images in the dataset class and to other aspects covered during the labs). (1 point)
You must also use at least 4 torchvision transformations (from which 2 have a random factor) that will be passed to the dataset class (search for examples in the given documentation).

Your training pipeline must run both on CPU and GPU, using a device parameter.
Upload your homework in the /Lab04/Solution directory. If you are doing your homework in a Jupyter Notebook, add the "Open in Colab" option.

Bonus:

- 1 point for a very efficient pipeline (this does not refer to common practices used in every online example). Must be proven by timing and A/B testing.
- 1 point for very modular and clear implementation (you can easily change the dataset, transformations, model and hyper-parameters and the training pipeline remains the same; the Dataset class implementation should be adjusted such that it is becomes a dataset wrapper, that wraps any other custom Dataset class that has to be written when changing datasets).