Machine Learning CW4

Objective. The goal of this coursework is to gain a deeper insight into the performances and applicability of different machine learning algorithms including random forests, multilayer Perceptrons (MLPs), and convolutional neural networks (CNNs). We will conduct an experimental study of these approaches on image classification tasks.

Datasets. We will use *CIFAR100* and *Fashion-MNIST* image classification benchmark datasets: The CIFAR100 dataset contains 60,000 color images of size 32×32 in 100 classes [5]. Among them, 50,000 images were originally allocated for training (500 images per class) while the remaining 10,000 images were reserved for testing. The *Fashion-MNIST* dataset provides 70,000 grayscale images of size 28×28 from 10 classes (60,000 training images and 10,000 testing images). *CIFAR100* and *Fashion-MNIST* can be downloaded from [4] and [6], respectively. The details of the datasets can also be found there.

For both datasets, select 20,000 images from the original training sets. Among these 20,000 images, use 10,000 images for training and the remaining 10,000 for validation: Any selection methods are allowed, e.g. for *CIFAR100*, one could randomly select 200 images from each class. For testing, please use the entire test sets (of size 10,000) for both datasets.

Task 1 (basic experiments and analysis). Please perform experiments with 1) random forests; 2) MLPs, and 3) CNNs on *CIFAR100* and *Fashion-MNIST* and analyze the corresponding results. Minimally, you need to provide **solid discussions** on

- classification accuracies and run-times (training and testing) of all algorithms on each dataset;
- details of hyperparameter selection, e.g. considered hyperparameters and the corresponding search ranges

accompanied by plots and/or tables supporting these findings.

Task 2 (advanced experiments and analysis). Make additional efforts: Examples include (but are not limited to)

- visualization of different hyperparameter combinations for each algorithm;
- accuracies and run-times of different algorithms with varying number of training images (e.g. $\{1{,}000, 2{,}000, \ldots, 10{,}000\}$ and/or $\{10{,}000, 20{,}000, \ldots, 40{,}000\}$);
- experiments with other classification algorithms (e.g. support vector machines and more than one type of CNNs);
- experiments on other datasets;
- ablation studies, e.g. analyzing the effect of removing and/or adding specific components in each algorithm;
- combination of different algorithms, e.g. convolution layers of CNNs for feature extraction combined with support vector machine classifiers.

What to hand on? Please submit a single zip file that contains your code package and a pdf document of your report. Please format the submission as in **'StudentID_Name.zip'**, e.g. 20221234_KwangInKim.zip. The accompanying report should also be formatted as in **'StudentID_Name.pdf'**. If your submission does not meet this file name requirement, the final mark will become 80% of the initial mark. Also, please type your report: If your submission is not typed (e.g. a scanned pdf of handwritten solutions), the final mark will be reduced to 50%.

1. ***Report.*** Please use NeurIPS2022 style files (available at `https://neurips.cc/Conferences/2022/PaperInformation/StyleFiles`) with 'final' option:

```
\usepackage[final]{neurips_2022}.
```

Your report should be maximum of six pages long including figures and tables but excluding the references and appendices. The report should **discuss** the main findings (e.g. comparisons of the classification accuracies and run-times of different algorithms) and support these discussions with the corresponding evidence (e.g. figures and tables summarizing the experimental results). It is expected that your discussions refer to figures and/or tables. Simply placing figures and tables in the report is not enough: **Any table or figure that is not accompanied by a discussion will be ignored during marking**. Similarly, if you present any material in the appendices, please make sure that they are referred to in the main text (within the six-page limit).

2. ***Code.*** **You do not have to implement the learning algorithms by yourself**: You can use existing code packages including the MLP code package that you submitted for CW3. However, note that you cannot use other UNIST students' code. For each algorithm and dataset, you need to submit a Python code that 1) reads training data for each dataset (*CIFAR100* and *Fashion-MNIST*); 2) selects 20,000 images (10,000 for training and the remaining 10,000 for validation) from the original training set for each dataset; 3) trains and tests the respective algorithm: Your code could call existing machine learning software libraries for training and testing. Please do not include the hyperparameter optimization step in the submitted code: Specify the final hyperparameters in the submitted code. Your code should meet the **requirements stated in the accompanying 'README.md' file**: The code submission should specify the python version and required packages. Please name your code 'RandomForestsCIFAR100.py', 'MLPCIFAR100.py', 'CNNCIFAR100.py', 'RandomForestsFashionMNIST.py', 'MLPFashionMNIST.py', and 'CNNFashionMNIST.py'.

Grading.

1. Code (20/100): We will simply check if your code runs as specified. If your code successfully implements the training and testing phases of different algorithms and it meets all requirements listed in the previous paragraph, you will earn full marks.

2. Report (80/100): 40 marks are assigned for Task 1 and Task 2, respectively. We will asses the report based on how convincingly the main findings are communicated: Please note that good discussions are important: You could characterize the behaviors of different algorithms, e.g. by providing confusion matrices or showing example images on which the algorithms made mistakes. Please see Section 4 of [1], Section 4 of [3], and Section 4 of [2] among others, for example discussions: Basically, your report is expected to be similar to Section 4's of these academic papers.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[2] Kwang In Kim, Juhyun Park, and James Tompkin. High-order tensor regularization with application to attribute ranking. In *CVPR*, pages 4334–4357, 2018.

[3] Kwang In Kim, James Tompkin, and Christian Richardt. Predictor combination at test time. In *ICCV*, pages 3553–3561, 2017.

[4] Alex Krizhevsky. CIFAR10 and CIFAR100 datasets, 2009. https://www.cs.toronto.edu/~kriz/cifar.html.

[5] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[6] Han Xiao, Kashif Rasul, and Roland Vollgraf. FashionMNIST: a novel image dataset for benchmarking machine learning algorithms. In *arXiv:1708.07747*, 2017. https://www.kaggle.com/datasets/zalando-research/fashionmnist.