COMP5328 - Advanced Machine Learning

Assignment 2

Due: 2 November 2018, 5:00PM

This assignment is to be completed in groups 2 to 3 students. It is worth 20% of your total mark.

Objective

The goal of this assignment is to study how to learn with label noise. Specifically, you need to use at least **two** methods to classify real world images with noisy labels into a set of categories. Then, you need to compare the performance of these classifiers and analyze the robustness of label noise methods.

The datasets are quite large, so you need to be smart on which methods you gonna use and perhaps perform a pre-processing step to reduce the amount of computation. Part of your marks will be a function of the performance of your classifier on the test set.

1 Dataset description

In this assignment, you need to apply label noise methods on two real world image databases: Fashion-MNIST[1] and CIFAR[2]. However, we re-organize and re-sample the datasets to construct new datasets with noisy labels for binary classification.

The constructed dataset can be downloaded from Canvas. Download it and put it into any folder you like. Here, we assume your folder name is "your_filepath". The datasets are stored as .npz files: "mnist_dataset.npz" and "cifar_dataset.npz", which can be loaded by using the following code:

```
import numpy as np
# assume that dataset is stored in the file 'dataset.npz'
```

[1] Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. Han Xiao, Kashif Rasul, Roland Vollgraf. arXiv:1708.07747.

[2] https://www.cs.toronto.edu/ kriz/cifar.html

```
dataset = np.load('your_filepath/dataset.npz')
Xtr = dataset['Xtr']
Str = dataset['Str']
Xts = dataset['Xts']
Yts = dataset['Yts']
```

1. Training features and labels:
   - Xtr: shape=$(10000, d)$. There are $10,000$ instances. The raw data are 28×28 (for Fashion-MNIST) or $32 \times 32 \times 3$ (for CIFAR) images, which are reshaped to features with dimension $d = 784$ or $d = 3072$.
   - Str: shape=$(10000, 1)$. There are $10,000$ noisy labels for the corresponding instances.
   - These $10,000$ instances belong to two categories. The corresponding labels for these two categories are 0 and 1. These training examples are with label noise. The flip rates are $\rho_0 = p(S = 1|Y = 0) = 0.2$ and $\rho_1 = p(S = 0|Y = 1) = 0.4$, where S and Y are the variables of noisy labels and true labels, respectively.
   - Note that do not use all the $10,000$ examples to train your models. You are required to independently and randomly sample $8,000$ examples from the $10,000$ examples to train every classifier. The reported performance of each model should be the average performance of at least 10 learned classifiers.
2. Test features and labels:
   - Xts: shape=$(2000, d)$. There are $2,000$ instances. Same with the training instances, each instance is represented by a feature with dimension $d = 784$ (for Fashion-MNIST) or $d = 3072$ (for CIFAR).
   - Yts: shape=$(2000, 1)$. There are $2,000$ true labels for the corresponding instances. The labels are also from the label set $\{0, 1\}$.

Note that the validation set is not provided, but you can randomly pick a subset from the training examples ($8,000$ sampled examples) for validation. You are suggested to use a validation set for selecting parameters such as learning rate and parameters of regularization term (if any) or for avoiding overfitting.

The flip rates ρ_0 and ρ_1 are given and you can use these parameters directly in your algorithm. But there are also 10 points for the estimation of the flip rates, which will be described in the following section. If you estimate the flip rate, you need only to report the estimated flip rates and need not to use them to train any another classifier.

The labels of the 2,000 test examples are given, you will analyse the performance of your methods by exploiting the 2,000 test examples. It is **NOT** allowed to use any examples from the test set for training or for tuning the parameters of your model; or it will be considered as cheating.

You can use the importance reweighting method (which is introduced in Tutorial 11) as one of your methods.

1.1 Performance Evaluation

The performance of each classifier will be evaluated in terms of the top-1 accuracy metric, i.e.

$$\text{accuracy} = \frac{\text{number of correctly classified examples}}{\text{total number of test examples}} * 100\%.$$

Note that, we expect you to have a rigorous performance evaluation. To provide an evaluation of the models, you are required to train each model at least 10 times. In each time, independently and randomly sample $8,000$ examples from the $10,000$ training examples and use the sampled $8,000$ examples as new training examples to train the model. Finally, use the average accuracy and the standard derivation on the test data to compare effectiveness and robustness.

2 Task description

Your task is to determine / implement at least two methods for the given datasets to classify real-world images into categories and write a report to analyze the effectiveness and robustness of each algorithm on the datasets. The score allocation is as follows:

- Report: max 80 points
- Code: max 20 points

Please see section 4 for the detailed marking scheme. The report and the code are to be submitted *Canvas* by the due date.

2.1 Report

The report should be organized similar to research papers, and should contain the following sections:

- In **abstract**, you should briefly introduce the topic of this assignment, your methods, and describe the organization of your report.

- In **introduction**, you should first introduce the problem of learning with label noise, and then its significances and applications. You should give an overview of the methods you want to use.

- In **related work**, you are expected to review the main idea of related label noise methods (including their advantages and disadvantages).

- In **methods**, you should describe the details of your methods, including the definition of cost functions, the theoretical foundations or views (if any) of designed cost function, and the optimization methods. If you estimate the flip rates, you should also describe the details of the estimation methods, include objective function, theoretical foundations (if any), and optimization algorithms.

- In **experiments**, you should introduce your experimental setup (e.g., datasets, algorithms, evaluation metric, etc.). Then, you should show the experimental results, compare, and analyze your results. If possible, give your personal reflection or thoughts on these results.

- In **conclusion**, you should summarize your methods, results, and your insights for the future work.

- In **references**, you should list all references cited in your report and formatted all references in a consistent way.

2.2 Programming language and libraries

This assignment must be submitted in Python3. You are allowed to use external libraries for optimisation and linear algebraic calculations. If you have any ambiguity whether you can use a particular library or a function, please post on canvas under the "Assignment 2" thread.

3 Instructions to hand in the assignment

1. Go to Canvas and upload the following files/folders compressed together as a zip file.

   (a) report (a pdf file)
   The report should include all member's details (student IDs and names).

(b) code (a folder)

i. algorithm (a sub-folder)
Your code (could be multiple files or a project)

ii. input data (a sub-folder)
Empty
Please do NOT include the dataset in the zip file as they are large.
We will copy the dataset to the input folder when we test the code.

Only one student needs to submit the zip file which must be named as student ID numbers of all group members separated by underscores. E.g. “xxxxxxxx_xxxxxxxx_xxxxxxxx.zip”.

2. Your submission should include the report and the code. A plagiarism checker will be used. Clearly provide instructions on how to run your code in the appendix of the report.

3. The report must clearly show (i) details of your classifier, (ii) the results from your classifier, including the predicted results from your classifier on test examples and accuracies, (iii) run-time, and (iv) hardware and software specifications of the computer that you used for performance evaluations.

4. We have provided a template for writing the report. Note that you have to strictly follow the format of the template. The maximum length of the report is 20 (including references).

5. A penalty of MINUS 20 percent points (−20%) per each day after the due date. Maximum delay is 5 (five) days, after that assignments will not be accepted.

6. Remember, the due date to submit them on Canvas is **2 November 2018, 5:00PM**.

4 Marking scheme

Category	Criterion	Marks	Comments
Report [80]	Abstract [3] •Problem, methods, and organization. Introduction [4] •What is the problem you intend to solve? •Why is this problem important? Previous work [10] •Previous relevant methods used in literature? Label noise methods [20] •Pre-processing (if any). •Label noise method's formulation. •Cross-validation method for model selection or avoiding overfitting (if any). Noise rate estimation methods [10] •Noise rate estimation method's formulation. •Experiments. Experiments and discussions [20] •Experiments, comparisons and evaluations. •Extensive analysis and discussion of results. •Relevant personal relection. Conclusions and future work [3] •Meaningful conclusions based on results. •Meaningful future work suggested. Presentation [5] •Academic style, grammatical sentences, no spelling mistakes. •Good structure and layout, consistent formatting. •Appropriate citation and referencing. •Use graphs and tables to summarize data.		

	Other [5] •At the discretion of the marker: for impressing the marker, excelling expectation, etc. Examples include fast code, using LaTeX, etc.		
Code [20]	•Code runs and classifies within a feasible time •Well organized, commented and documented		
Penalties [−]	•Badly written code: [−20] •Not including instructions on how to run your code: [−20] •Late submission: [−20% per day]		

Note: Marks for each category is indicated in square brackets. The minimum mark for the assignment will be 0 (zero).