Lab Assignment 4

Machine Learning Due on: 24th April 2023 11:59 PM | M.M.:100 CS 503

Instructions:

1. The research assignment is for those students who are not doing the project.
2. This lab assignment needs to be done individually, and it carries 13% weightage of your total lab requirements.
3. Start by going through the research paper first before you start coding.
4. For queries, you should comment on Google Classroom. Alternatively, you may email the course TAs or course instructors.
5. All code must be written in python using a jupyter/colab notebook.
6. Submission is only through Google Classroom. The code, as well as the accompanying observations must be included in the same jupyter/colab notebook separately for each task.
7. Code Readability is very important. Modularize your code using classes and functions that can be flexibly reused wherever necessary. Also, use self-explanatory variable names and add comments to describe your approach wherever necessary.
8. **Students are expected to follow the honour code of the class**. Discussions and interactions with your classmates to take help in this assignment are not allowed.

This research will make you explore a machine learning paradigm called *Federated learning* (FL). FL aims to learn a global machine-learning model collaboratively from data generated by and residing on several remote devices or clients locally. FL stands to produce highly accurate statistical models by aggregating knowledge from disparate data sources while preserving data privacy. The general architecture of FL is shown in Fig 1.

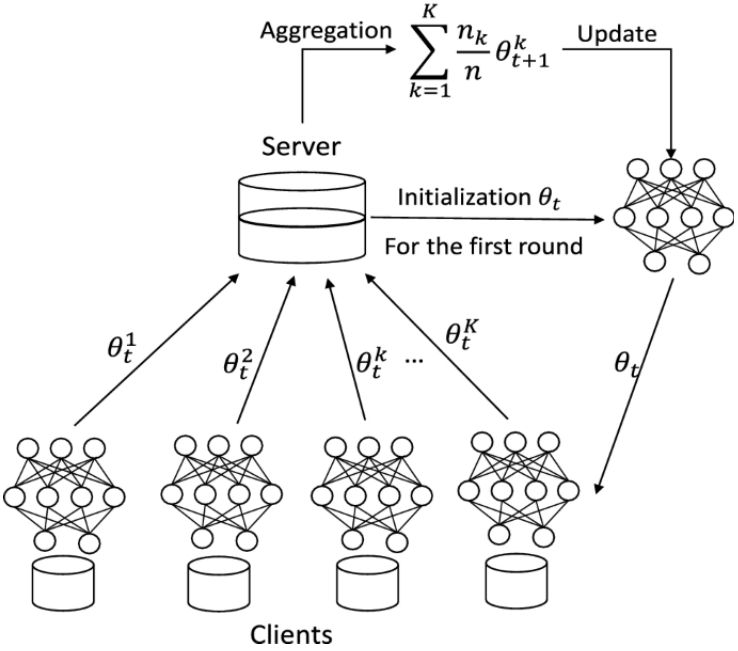
However, the resulting systems should be accurate and satisfy several pragmatic constraints such as fairness, robustness, and privacy. Simultaneously satisfying these varied constraints can be exceptionally difficult. For this assignment, we will be focusing on the following research paper by Li et al.[2]:

`T. Li, S. Hu, A. Beirami, and V. Smith. Ditto: Fair and robust federated learning through personalization. In International Conference on Machine Learning, pages 6357-6368. PMLR, 2021`

The work focuses on issues of accuracy, fairness, and robustness in FL. The authors investigate a simple scalable technique that satisfies all the above three constraints. They identify statistical heterogeneity as the root cause for tension between these constraints which in itself is a key to the solution. In particular, they suggest that methods for personalized FL[1] may provide inherent benefits in terms of fairness and robustness.

You are advised to go through the above-mentioned research paper in order to complete this assignment. Below, there are definitions of the three important terms w.r.t. FL.

[1] which adapts to the heterogeneity in federated settings by learning distinct models for each device



Architecture for the working of FL system. (Source: Google Images)

Definitions:

Definition 1 (Robustness in FL [2]). *Robustness mainly concerns **byzantine robustness** wherein malicious devices can send arbitrary updates to the server to compromise training. It can be in the form of label poisoning (sending flipped class labels), random updates (sending random zero Gaussian updates), or model replacement (scaling update values).*

Definition 2 (Fairness in FL [2]). *A model w_1 is fairer than w_2 if the test performance distribution of w_1 across the network (devices) is more uniform than that of w_2, i.e., $std\{F_k(w_1)\}_{k\in[K]} < std\{F_k(w_2)\}_{k\in[K]}$ where $F_k(\cdot)$ denotes the test loss on device $k \in [K]$, and $std\{\cdot\}$ denotes the standard deviation.*

A lower standard deviation implies lesser variation in loss values across different devices. Such fairness is known as **representation disparity**.

Definition 3 (Personalization in FL). *Personalization is a natural approach to improve accuracy. In a federated setting, it is the ability to customize the global model to each device's individual preferences or characteristics. One way to achieve personalization in federated learning is by using fine-tuning the global model's parameters to match the characteristics of each device or user, i.e., further training them locally.*

Ditto (present paper) also tries to achieve personalization via a different routine (refer to Algorithm 1 in the paper [2]).

Lab Assignment 4

Machine Learning Due on: 24th April 2023 11:59 PM | M.M.:100 CS 503

Code:

The code for the paper is available at https://github.com/litian96/ditto. You are free to use the author's code along with any other library/API that can help you to complete the tasks in this assignment.

You are required to perform the following tasks in the current assignment.

Task 1 :

The first task is to understand the Ditto paper and reproduce its results. In this task, you must reproduce the robustness (Section 4.1), fairness (Sections 4.2, & 4.3), and accuracy (Section 4.4) results on the **Fashion MNIST (FMNIST)** and **FEMNIST (skewed) Datasets** provided in Appendix C of the paper. Use the value of lambda given in the paper for these datasets. In the personalization task (e.g., Table 2 of the paper), you can execute any other two baselines (additional to Ditto), say L2SGD and EWC. The code of these baselines is available in the same GitHub repo.

Prepare a report with the tables and figures of reproduced results. If you feel there were some challenges in reproducing certain results, report them.

(Note that it is quite possible that you might be unable to replicate certain results; in such cases, report your results and related observations regarding the possible reasons behind it). Leave Table 3 and Figure 6.

[30 Marks]

Task 2 :

The DITTO algorithm tries to achieve personalization using Algorithm 1. Let's develop another possible algorithm (say FedAvgPer) that works as follows:

1. After a global model is learned in simple vanilla federated averaging[2]. The final global model is shared with the clients.
2. Now, clients will directly run a few local epochs on their own data and update model parameters. This will add each client's own personal touch. (We are not adding any constraints of the combined loss objective and regularization between the local and global models as in Ditto).
3. Now, after updating model parameters locally, each client has its own personalized FL model.

[2] https://arxiv.org/pdf/1602.05629.pdf

Compare FedAvgPer with Ditto and other existing baselines in Task 1, i.e., reproduce all the results gathered in the previous task with this updated approach. Is FedAvgPer better than Ditto ? Report your observations.

You can club the results of Task 1 and Task 2 in the same graphs and Tables).

[20 Marks]

Task 3 :

Now consider another dataset called the ADULT dataset available at https://archive.ics.uci.edu/ml/datasets/adult. It is a census dataset usually used to predict whether income exceeds $50K$ per year. Let us consider a scenario where we have clients in different countries/regions, each generating their own data. To mimic this scenario, divide the complete dataset into clients based on column 'native-country'. Further, consider a fairness metric called Equal Opportunity[3]. Equalized opportunity means matching the true positive rates (TPR) for different values of a protected/sensitive attribute such as for ADULT dataset matching TPR for income prediction for different values of gender. If for both MALE and FEMALE, the TPR rates are same then the model is fair otherwise its unfair.

Here, the fairness criteria has been evaluated based on the sensitive attribute column 'gender' in the dataset.

Your task is to create a new loss function considering the combination of accuracy and Equal Opportunity fairness metrics (considering gender) for improving personalization and fairness performance. Report the different results for this approach.

(Note that you can use the base model as SVM for the adult dataset instead of CNN used in previous parts for FMNIST, FEMINST image datasets).

[25 Marks]

Task 4 :

In real-world decision-making systems, classification models must not only be accurate but also indicate when they are likely to be incorrect. Specifically, a model should provide a calibrated confidence measure in addition to its prediction. In other words, the probability associated with the predicted class label should reflect its ground truth correctness likelihood. Good confidence estimates for the predicted label provide a valuable extra bit of information to establish trustworthiness with the user – especially for neural networks, whose classification decisions are often difficult to interpret.

[3] https://ocw.mit.edu/courses/res-ec-001-exploring-fairness-in-machine-learning-for-international-deve
pages/module-three-framework/fairness-criteria/

A metric to analyze the level of calibration of a model is Maximum Calibration Error (MCE) (refer to Equation 4 in paper [1]).

Your task is to report the MCE for different personalized models obtained using DITTO (Algorithm 1 in [2]).

Further, calibrate the local client models using the Platt scaling method in the paper [1] (Section 4.1) and again compare the calibration errors before and after applying Platt scaling calibration. Is there any advantage of calibration for the clients?

[25 Marks]

References

[1] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On calibration of modern neural networks. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1321–1330. PMLR, 06–11 Aug 2017. https://arxiv.org/pdf/1706.04599.pdf.

[2] T. Li, S. Hu, A. Beirami, and V. Smith. Ditto: Fair and robust federated learning through personalization. In *International Conference on Machine Learning*, pages 6357–6368. PMLR, 2021. https://arxiv.org/pdf/2012.04221.pdf.