PriDe: Privacy Preserving Federated Learning on a Deep Neural Network

Sean Pavlak
Case Western Reserve University
Cleveland, OH
seanpavlak@gmail.com

ABSTRACT

Groundbreaking developments in medicine, economics, etc. are taking place everyday due to the field of machine learning. However, rarely is your data truly protected. The modern paradigm in data science and machine learning is to aggregate everyones sensitive data in a data center and protect it behind a layer of encryption. This, however, has been shown to be an unsafe practice; as shown in major data leaks by Apple, Equifax, and many more. A potential solution is to allow users to retain their data locally and protect it locally, while still retaining the ability for learning models to be developed using this data. Federated Learning is a recent advance in privacy protection and is a real solution to this issue. This technology is a means to learn a shared model by aggregating locally-computed updates while the sensitive training data distributed locally across many parties remains local to each party. I intend on implementing federated learning of deep networks based on iterative model averaging on a chosen data set, MNIST data set. However, this learning method is vulnerable to differential attacks, which could originate from any party contributing during federated optimization. In such an attack, a clientâĂŹs contribution during training and information about their data set is revealed through analyzing the distributed model. I plan on protecting against this by implementing a set of privacy preserving measures to ensure that this sensitive data is protected against adversarial parties during training. The aim is also to balance the trade-off between privacy loss and model performance.

General Terms

Federated Learning, Federated Optimization, Neural Network, Privacy

1. INTRODUCTION

In this project I set out to train a deep neural network using the MNIST data set distributed between multiple parties while maintaining privacy for each user. The primary method I intend on using is federated learning. This method of machine learning is designed to act on data stores that are held locally across many devices, parties, which is perfect for my task.

Federated learning inherently is more secure than standard traditional learning methods as it does not require centralizing data on one machine or in a data center. This method allows many parties to collaboratively learn a shared prediction model while keeping all the training data on device. This learning method was created by Google to train models on data stored locally across android mobile devices. It works by first downloading the current model, improves it by learning from local data, and then summarizes the changes as a small focused update. This update to the model is then sent to a third party in an encrypted form and is averaged with other user updates to improve the shared model [1].

According to H. B. McMahan, et al. federated learning is designed to train from real world data on mobile devices, train on sensitive or large data sets, and to perform supervised learning on data that can be inferred naturally from user interaction [6]. A great example of this is image classification, as it is real world data that can contain sensitive information and can be easily labeled with slight user interaction.

Skin cancer image classification is a significant problem being solved with deep learning and convolutional neural networks [2]. Researchers at Stanford were presented with a data set of 130,000 clinical images. This enabled them to create an image analysis application running in a mobile environment that can classify skin lesions to hopefully predict skin cancer from a single image. This technology is undoubtedly significant, however; using a data set presents researchers with a set of issues. Generally limited access to secure data preventing the learning model from increasing in accuracy over time.

This is an excellent example of image classification that solves a problem plaguing millions of people worldwide that is only as accurate as the training set allows it to be. By training with federated learning you would be able to continually train the model as users use the model, and overtime potentially achieve the theoretical limit of predicting and classifying different forms of skin cancer. All while preventing the user's data from ever leaving their phone and maintaining a high level of privacy for a data sensitive image.

While there is a natural predisposition to privacy with federated learning, due to a user's data never being sent from device. Yet it is still incredibly necessary to encrypt the model update prior to sending it from the users device to the trusted third party. In the event that an adversary is able to access the current state of the model and access the optimization update sent from a party's device, it is possible that the adversary would be able to infer the sensitive local data without ever having access to it. In order to protect against this the update needs to be encrypted prior to being uploaded into the third party's possession.

Utilizing a form of differential privacy or homomorphic encryption has been utilized for some studies with success, lending me to believe that this level of privacy should be sufficient in preventing adversaries.

2. METHODS

For this project I intend on performing it in a series of iterations.

Firstly, I plan on developing a simple form of federated learning by training on only two parties, sets of data, at first. The intent behind this is that a multiple party system, or even an n-party system, can very quickly get out of hand and is very difficult to maintain while a lot of systems are not ironed out and developed. In the simple form I also will train on a small set of a simple data set, such as the Iris data set.

Training this classifier will ideally be much quicker, yielding to faster build times and consequently a quicker development time.

In the later half of this development cycle I will also implement an encryption system for the model weights, so they are secure. My intent is to utilize a python library, such as PHE, to easily encrypt the data.

After I have this simple federated learning method completely functional, I will begin expanding the simple method. I will need to extend the method to accept MNIST data rather than Iris data and it will need to be expanded into an n-party system.

I will also ideally be able to optimize the neural network to perform as quickly as possible while still maintaining accuracy. It is also important that I extend the level of encryption in order to secure user data as well as possible.

3. NEXT STEPS

I was able to develop a simple form of federated learning using two parties and training on a simple training set, Iris flower data set. The neural network implemented contains a hidden layer utilizing a sigmoid function along with a set of activation layers I was also able to encrypt the model weights using Partially Homomorphic Encryption, using the PHE python library.

In the next few weeks I plan on extending the program to be able to accept and classify data from the MNIST data set. Ideally I would also like to extend my federated learning from a two party system to a 3 party system, and potentially even an n-party system. It may also be necessary to optimize the neural network to perform more efficiently.

I have also began working of training a convolutional neural network using KERAS on the MNIST data set, achieving a high accuracy. I also intend on extending, and simplifying, my code using KERAS.

Also currently I am encrypting the model weights using partially homomorphic encryption from the PHE python library. I would like to test multiple other encryption methods in order to try and determine a balance between privacy and training time.

4. EXPECTED RESULTS

From my current state, I am able to train Iris data from two separate partied on a federated learning neural network with an accuracy of 96.7%. The current training time is roughly 50 seconds, while only using 60 data points as training data. This accuracy is notable, however; in order for federated learning to be useful it must be able to develop a training model update in a relatively small amount of time.

With the next steps I believe I will be able to maintain a similar accuracy, and from prior work I have been able to develop a deep neural network on the MNIST data set achieving 98% accuracy. This was done using KERAS and a number of hidden layers along with multiple activation layers. I believe that I will be able to utilize KERAS and drop the training time down significantly.

According to R. C. Geyer et al. "Empirical studies suggest that given a sufficiently large number of participating clients, our proposed procedure can maintain client-level differential privacy at only a minor cost in model performance." [3] This study gives me reason to believe that as the number of parties increases I will be able maintain a high accuracy while keeping individual training times reasonable.

[1, 3, 4, 5, 6, 7, 8]

5. REFERENCES

[1] Federated learning: Collaborative machine learning without centralized training data, Apr 2017.

[2] R. A. N. J. K. S. M. S. H. M. B. . S. T. Andre Esteva, Brett Kuprel. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017.

[3] R. C. Geyer, T. Klein, and M. Nabi. Differentially Private Federated Learning: A Client Level Perspective. *ArXiv e-prints*, Dec. 2017.

[4] T. Klein. Differentially private federated learning: A client level perspective, Jan 2018.

[5] J. Konecný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon. Federated learning: Strategies for improving communication efficiency. *CoRR*, abs/1610.05492, 2016.

[6] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas. Federated learning of deep networks using model averaging. *CoRR*, abs/1602.05629, 2016.

[7] M. Nabi. Privacy-preserving collaborative machine learning âĂŞ sap leonardo machine learning research âĂŞ medium, Sep 2017.

[8] J. Rodriguez. What's new in deep learning research: Understanding federated learning, Feb 2018.