Privacy Preserving Data Classification Using Inner Product Encryption

Damien Ligier[1(✉)], Sergiu Carpov[1], Caroline Fontaine[2], and Renaud Sirdey[1]

[1] CEA LIST, Point Courrier 172, 91191 Gif-sur-Yvette Cedex, France
{damien.ligier,sergiu.carpov,renaud.sirdey}@cea.fr
[2] CNRS/Lab-STICC and Telecom Bretagne and UEB,
Technopôle Brest-Iroise, France
caroline.fontaine@telecom-bretagne.eu

Abstract. In the context of data outsourcing more and more concerns raise about the privacy of user's data. One solution is to outsource the data in encrypted form. Meanwhile obtaining a service based on machine learning predictions on user data remains very important in real-life situations.

This paper presents ways to combine machine learning algorithms and IPE in order to perform classification on encrypted data. The proposed privacy preserving classification schemes allow to keep user's data encrypted but at the same time revealing to a server classification results on this data. We study the performance of such classification schemes and their information leakage.

Keywords: Functional encryption · Inner-product encryption · Classification · Linear classification

1 Introduction

With the generalization of data outsourcing, more and more concerns raise about the privacy and the security of the outsourced data. In this context, machine learning methods have to be conceived and deployed by keeping in mind and assuring the user's privacy.

In a privacy preserving data classification process, one has to be able to extract knowledge (e.g. in the case of a classifier, deduct the class label of an individual without compromising his private data) by assuring the protection of the sensitive data and, if possible, by hiding data access patterns from which useful properties could be inferred.

In this work we propose a privacy preserving classification algorithm based on functional encryption, in particular the inner product encryption. The performance of the classification algorithm is evaluated on the MNIST database [3].

An inner product encryption scheme is a functional encryption one that enables the evaluation of inner products. In those public encryption schemes vectors are encrypted and each secret key is associated with one vector.

R. Deng et al. (Eds.): SecureComm 2016, LNICST 198, pp. 755–757, 2017.
DOI: 10.1007/978-3-319-59608-2_44

For example if c_v is an encryption of the vector v and sk_w is a secret key associated with the vector w, when one decrypts c_v with sk_w he gets $\langle v, w \rangle$. Note that secret keys are generated with the master secret key by the authority.

In the use case we focus in the paper, there is an entity called *server* that has performed a training step of a linear classifier. Thus he has a set of linear classification coefficients and he wants to keep them secret. There are many *users* that have informations that they want to keep secret as well but they also want to release classification results to the server (for example for obtaining a service). We introduce a third party that both of the server and the users can trust and we call it *authority*. His goal is in a first time to check that the server's coefficients are not dishonest and in a second time, to generate an instance of an inner product encryption to perform the classification over the encrypted inputs.

2 Privacy Preserving Classification

A linear classification algorithm makes a decision on the membership of an input data object, based on a linear combination of its features (characteristics). For example, in an image classification algorithm the input object is an image and the features can be image pixels. In a binary classification, the decision is made as a function of a threshold overrun by the dot product between object features and linear classifier coefficients.

In this work we propose a privacy preserving data classification method. Input data is encrypted using an inner product encryption scheme. In the context of ML algorithms, the inner product encryption can be seen as a linear binary classifier. In order to perform a multi-class linear classification, we need to compute several inner products on the same input data. Usually, linear classifiers provide worse results when compared to other more elaborate classification methods. At the same time, only a linear classifier is able to provide a prediction for data encrypted using the inner product encryption.

In order to fill this gap we propose a combined classification method, in which a linear classifier is applied to encrypted data and is followed by a more complex classification algorithm (for example an ensemble method in our case but not limited to). For each piece of input encrypted data, several inner products are computed. These products are then used as input features for a second, more elaborate, classifier. In this way, we are able to perform classification of encrypted data with increased performance in terms of an evaluation metric (e.g. error rate).

3 Performance

We use our implementation (in C++ using FLINT library [2]) of the functional encryption for inner product scheme of Agrawal *et al.* [1] which provides full security under the DDH assumption. We work in a group $\mathbb{F}_p^*$ such that p is a safe prime of approximatively 2048-bits.

We try our construction with the MNIST database [3] which is a collection of handwritten digit images (28×28 pixels with 256 levels of grey). So in this use case, the classifier has 10 output classes (digits from 0 to 9).

The experiments were performed on a regular laptop computer with an Intel Core i7-4650U CPU and 8 GB of RAM. A plaintext and a secret key have size about one kB and a ciphertext has a size about 200 kB.

The algorithm to generate a secret key associated with a vector and the encryption algorithm take less than a second to be computed. The decryption algorithm takes about 23 s to be computed.

We get 14% of error rate with a single linear classifier, and 7% using a second classifier after the first one which does not take significant time to perform.

4 Classification Security

We emphasize that the use case of a such construction can be unsecure even if the cryptographic scheme is secure. An attacker gets a system of diophantine equations. The easiness of solving it and the precision of the description of the inputs increase the ability to compute which vector has been encrypted.

5 Conclusion and Future Work

In this work we have used an instantiation of an inner product encryption in order to perform classification over encrypted data. The learning process is kept secret as only linear classifiers coefficients are public. In the use-case we study, we have a trusted authority, a server computing the classification and the users who encrypt their data. Obtained execution times are reasonably small (a prediction is made in approximatively 69 s without any parallelization) as well as the size of the ciphertexts. We have studied a method to ensure that we cannot find original image from the inner product values. In perspective, we consider to study more deeply the leakage of inner product encryption schemes and to propose methods to lower it.

References

1. Agrawal, S., Libert, B., Stehle, D.: Fully Secure Functional Encryption for Inner Products, from Standard Assumptions (2015)
2. Hart, W., Johansson, F., Pancratz, S.: FLINT: Fast Library for Number Theory, Version 2.4.0 (2013). http://flintlib.org
3. LeCun, Y., Cortes, C., Burges, C.J.: The MNIST Database. http://yann.lecun.com/exdb/mnist/