Encrypted Linear Regression using CKKS

Roberto Carboni, Roland Hostettler, Anders Ahlén, Subhrakanti Dey
Department of Electrical Engineering; Signals and Systems, Uppsala University, Uppsala, Sweden
Email: {roberto.carboni, roland.hostettler, anders.ahlen, subhrakanti.dey}@angstrom.uu.se

***Abstract*—In this paper, we consider the solution of encrypted linear regression using Homomorphic Encryption. This method allows the computation of linear regression in an encrypted environment. The proposed method consists of an iterative method based on a modified Goldschmidt sequence. Numerical results on synthetic data show that the method converges with minimal accuracy loss due to encryption noise, indicating that our approach is well-suited for homomorphically encrypted linear regression.**

I. Background

A. Introduction

Linear regression is a simplistic yet powerful and principled statistical approach to model the relationship between variables. However, performing linear regression over large amounts of data might be expensive in terms of computational cost. To overcome this problem, cloud computing can be used, which involves using shared servers managed by third-party providers. Since the data used in the cloud might be sensitive, it raises the need to provide a privacy-preserving environment. This can be achieved with Homomorphic Encryption (HE), which allows computations on encrypted data without decrypting it [1]. In this way, it is possible to share data with third-party providers while maintaining privacy. However, HE schemes, such as the Cheon–Kim–Kim–Song (CKKS) scheme [2], have strict constraints, such as the types and number of operations that can be computed on encrypted data. Hence, great care needs to be taken when implementing algorithms using HE. One of the most significant obstacles is the number of sequential operations that can be computed. Typically, HE schemes only support addition and/or (a limited number of) multiplications [3]. Hence, since divisions, and in particular matrix inverses are not supported, one has to resort to iterative algorithms to solve problems such as linear regression. However, classical iterative methods, such as the Gauss–Seidel method [4] require a large number of iterations.

Linear regression using HE has been considered in [5] and [6]. However, these methods do not compute all the operations in an HE environment. In particular, the methods rely on plaintext client-side matrix inversion. Furthermore, in [7], linear regression is implemented using the Paillier encryption scheme [8], which supports only addition and scalar multiplication. Computing dot products thus requires additional tools like additive secret sharing, secure fixed-point arithmetic, and interactive protocols. Since Paillier operates on integers, using real numbers directly can lead to precision loss.

In contrast, our approach utilizes the CKKS encryption scheme which supports both addition and multiplication on real or complex numbers. This allows the computation of linear regression in an encrypted environment and does not require additional techniques for the computation of matrix multiplications.

B. Linear Regression

Linear regression is a model that describes a linear relationship between a dependent (noisy) variable $\mathbf{y} \in \mathbb{R}^{d_y}$ and an independent variable (parameters) $\mathbf{x} \in \mathbb{R}^{d_x}$, according to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{r},$$

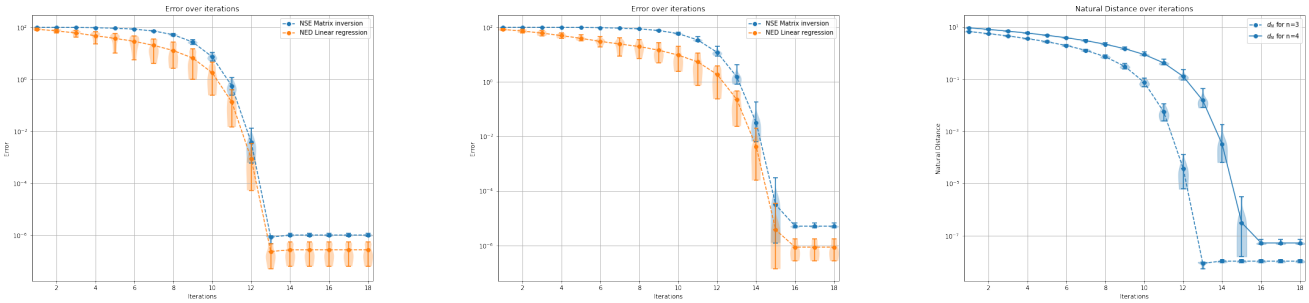
where $\mathbf{H} \in \mathbb{R}^{d_y \times d_x}$ is the observation matrix and $\mathbf{r} \in \mathbb{R}^{d_y}$ is a noise term. The observation matrix is obtained starting from a dataset D defined as a collection of input-output pairs $\{(\mathbf{u}_i, y_i)\}_{i=1}^{N}$ where $\mathbf{u}_i \in \mathbb{R}^{d_x}$ are the regressors (inputs) and $y_i \in \mathbb{R}$ are the corresponding target vectors (outputs) for each sample i. The observation matrix is then constructed as $\mathbf{H} = [\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_N]^T$. One way to estimate the parameters of this model is to use the least squares method [4], which has the analytical solution

$$\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}. \tag{1}$$

As shown in (1), the core of this method is to invert the matrix $\mathbf{A} = \mathbf{H}^T\mathbf{H}$ and compute the product $\mathbf{b} = \mathbf{H}^T\mathbf{y}$. Typical approaches for matrix inversion can be computed using one of the classical iterative methods such as Gauss elimination or Cholesky decomposition, see [4].

C. Homomorphic Encryption

In this work, in order to compute linear regression in an encrypted environment, we use HE [9], a kind of encryption that allows computations on encrypted data without decrypting it. The core of this type of encryption is the use of a structure-preserving map that ensures certain operations on plaintexts can be mirrored by operations on ciphertexts such that $\mathrm{encr}(a \diamond b) = \mathrm{encr}(a) \diamond \mathrm{encr}(b)$ and $\mathrm{decr}(\mathrm{encr}(a \diamond b)) = \mathrm{decr}(\mathrm{encr}(a) \diamond \mathrm{encr}(b)) = a \diamond b$, where encr and decr are the encryption and decryption functions respectively, and the symbol $\diamond$ represents any given operation. In this work we use the CKKS scheme [2]. This HE scheme is designed for approximate arithmetic with real or complex numbers and supports additions and a limited number of multiplications, making it a *leveled* HE scheme [10], which means that each sequential multiplication *consumes* one level, and the number of available levels is set by specific parameters before encryption. Moreover, the scheme is approximate due to the way noise is added during encryption, such that the result of an



(a) (b) (c)

Figure 1: Comparison of the errors for the two cases using synthetic data. (a) NSE and NED for the case with $n = 3$; (b) NSE and NED for the case with $n = 4$; (c) d_N for both cases.

encrypted operation will not be exact, but approximate, that is, $\mathrm{decr}(\mathrm{encr}(a) \diamond \mathrm{encr}(b)) \approx a \diamond b$. Moreover, the CKKS scheme allows us only to compute additions and multiplications.

II. Method

To solve (1) one can use iterative numerical methods, such as Gauss elimination, Cholesky decomposition, or gradient descent. However, these methods typically require a relatively large number of multiplications and divisions, rendering them not feasible to use together with CKKS. For these reasons, implementing linear regression in the HE environment is challenging. Instead, we propose the use of a modified Goldschmidt algorithm, first proposed in [11]. The proposed method solves the system (1) by numerically finding the inverse $\mathbf{A}^{-1}$. In particular, our approach is based on the modified Goldschmidt's algorithm proposed in [11]. Note that in our method all the operations are HE-friendly, meaning that they can be easily computed in an HE environment.

III. Results

We evaluate the proposed method on synthetic data. The error of the matrix inversion is evaluated using the Norm Spectral Error (NSE) and the Natural Distance (d_N) [12], while for the parameter estimation, we used the Normalized Euclidean Distance (NED).

To test our method, we consider two cases. For each case we generate 10 synthetic samples and execute our algorithm. We then compute the errors using the NSE, d_N, and NED defined above, followed by averaging these errors. For these tests, we used random vectors $\mathbf{v}$ with dimension $d_y = 100$ and $n = 3$ and $n = 4$. From these results reported in Figure 1, we can see that the method converges.

IV. Conclusions

In this work, we have defined a structured methodology for evaluating fully Homomorphically Encrypted linear regression. The proposed method allows the evaluation of a linear regression model where the input data remains encrypted throughout the entire process. We then evaluated the proposed method with synthetic data. Each computational step is performed using encrypted operations without requiring data decryption at any stage and without multiparty computation. The results from these experiments show that the proposed approach is capable of performing linear regression while preserving data privacy.

References

[1] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 169–178.

[2] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I 23*. Springer, 2017, pp. 409–437.

[3] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. H. Fitzek, and N. Aaraj, "Survey on fully homomorphic encryption, theory, and applications," *Proceedings of the IEEE*, vol. 110, no. 10, pp. 1572–1609, 2022.

[4] Å. Björck *et al.*, *Numerical methods in matrix computations*. Springer, 2015, vol. 59.

[5] B. Chen and X. Zheng, "Implementing linear regression with homomorphic encryption," *Procedia Computer Science*, vol. 202, pp. 324–329, 2022.

[6] I. Giacomelli, S. Jha, M. Joye, C. D. Page, and K. Yoon, "Privacy-preserving ridge regression with only linearly-homomorphic encryption," in *Applied Cryptography and Network Security: 16th International Conference, ACNS 2018, Leuven, Belgium, July 2-4, 2018, Proceedings 16*. Springer, 2018, pp. 243–261.

[7] R. Hall, S. E. Fienberg, and Y. Nardi, "Secure multiple linear regression based on homomorphic encryption," *Journal of official statistics*, vol. 27, no. 4, pp. 669–691, 2011.

[8] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International conference on the theory and applications of cryptographic techniques*. Springer, 1999, pp. 223–238.

[9] R. L. Rivest, L. Adleman, M. L. Dertouzos *et al.*, "On data banks and privacy homomorphisms," *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.

[10] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, "(Leveled) Fully Homomorphic Encryption Without Bootstrapping," *ACM Transactions on Computation Theory (TOCT)*, vol. 6, no. 3, pp. 1–36, 2014.

[11] T. M. Ahn, K. H. Lee, J. S. Yoo, and J. W. Yoon, "Cheap and fast iterative matrix inverse in encrypted domain," in *European Symposium on Research in Computer Security*. Springer, 2023, pp. 334–352.

[12] M. L. Nordenvaad and L. Svensson, "A map based estimator for inverse complex covariance matricies," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 3369–3372.