Using homomorphic encryption for linear regression on private data

Master’s Thesis

Machine learning is getting more widespread in the industry and especially in medical applications. While a large amount of data is required for training these machine learning models, especially in the latter scenario training data is hard to get, as it has a lot of requirements about data privacy, as sensitive information like genetic data is processed.

One approach to partially mitigate this issue is to use homomorphic encryption, encrypting the training data and machine learning model, performing the training on the ciphertexts and then decrypting only the final trained model. This thesis aims to apply this approach to the linear regression problem using an additively homomorphic encryption scheme and the gradient descent algorithm. The goal of this thesis is to find out how efficiently this can be done and what security properties can be achieved with this approach.

Scope of the work

- Implementing a linear regression on ciphertexts using gradient descent.
- An analysis on the limits of additively homomorphic encryption schemes for machine learning.
- Analyzing the security properties this approach achieves.
- Comparing the implementation with a classical linear regression approach regarding runtime performance, required resources, model accuracy and security.

Requirements

Following prior knowledge or skills are useful (or have to be acquired) for the master’s thesis:

- Familiarity with basic cryptography and homomorphic encryption.
- A basic understanding on machine learning and the gradient descent algorithm.
- Programming skills (preferably in Java, Python or Kotlin).
- Interest in the topic is strongly recommended.

Contact

In case of interest or for further information, please contact Robin Berger, robin.berger@kit.edu