Lecture: Pervasive Computing WS 2020/21

Student: Corina Laßnig

Topic: Privacy-Preserving Machine Learning: Threats and Solutions

Summary

Search queries, browsing history, purchase transactions, the videos and movies we watch are only a few examples of information that are being collected and stored on daily basis. This data collection happens within our mobile devices and computers, on the streets, and also in our own offices and homes. Such private data is being used for a variety of machine learning applications.

What is Machine Learning?

The main goal is to allow the computers to learn by self-acting without human intervention or assistance and adapt actions appropriate. A widely used definition is that of CMU † Professor Tom Mitchell:

"A computer program is said to learn from experience E, with respect to some class of tasks T and performance measure P if its performance at tasks in T as measured by P improves with experience E."

1. Supervised Learning:

Supervised Learning use labeled data where each feature vector corresponds with an output value. The output value could be a class label (classification) or a continuous value (regression). This labeled data is used to build models, this is the training phase, that can forecast new labels of feature vectors, this is the testing phase.

2. Unsupervised Learning:

With this type of learning, data is not labeled as feature vectors do not come with a class label or a response variable. The target in this case would be to find structure in the data. Clustering is probably the most common unsupervised learning technique, and its aim is to group a set of samples into different clusters. Samples in the same cluster are supposed to be relatively similar to each other, and different from samples in other clusters. Clustering means grouping a set of samples into a number of clusters.

3. Semi-supervised Learning:

Labeling data could be expensive and requires human experts or special devices. Only some of data gets labeled sometimes, while the vast majority persist unlabeled. Researchers found that even having a small part of labeled data can appreciably improve the learning process.

Threats and Solutions

In such systems, the data owner(s) send their data to the computation party that performs the required Machine Learning task and delivers the output to the results party. Such output could be a Machine Learning model that the results party can utilize for testing new samples. In other cases, the computation party might keep the Machine Learning model, and performs the task of testing new samples submitted by the results party and returning the testing results to the results party.

If all three roles are assumed by the same entity, then privacy is naturally preserved; however, when these roles are distributed across two or more entities, then privacy improving technologies are needed.

There are **multiple levels of threats** depending on the privacy leaks associated with the data sharing process. Examples of different threats are:

- Private Data in the Clear
- Reconstruction Attacks
- Model Inversion Attacks
- Membership Inference Attacks
- De-anonymization (Re-Identification)

Many **privacy-enhancing techniques** concentrated on allowing multiple input parties to collaboratively train ML models without releasing their private data in its original form.

Examples for privacy-enhancing techniques:

- Cryptographic Approaches
  - Homomorphic Encryption
  - Garbled Circuits
  - Secret Sharing
  - Secure Processors
- Perturbation Approaches
  - Differential Privacy (DP)
  - Local Differential Privacy
  - Dimensionality Reduction

Literatur:

1. Al-Rubaie, Mohammad and Chang, J. Morris (2018) 'Privacy Preserving Machine Learning: Threats and Solutions'. IEEE Security&Privacy

2. Xuan-Son Vu (2019) 'Privacy-Awareness in the Era of Big Data and Machine Learning'. Department of Computing Science

3. Al-Rubaie, Mohammad and Chang, J. Morris (2016) 'Reconstruction Attacks Against Mobile-Based Continuous Authentication Systems in the Cloud'. IEEE Transactions on Information Forensics and Security

4. Narayanan, Arvind and Shmatikov, Vitaly (2008) 'Robust de-anonymization of large sparse datasets'. Proceedings - IEEE Symposium on Security and Privacy

5. Bishop, Christopher M. (2007) 'Pattern Recognition And Machine Learning'.

6. Jianjiang Feng and Jain, A K (2011) 'Fingerprint Reconstruction: From Minutiae to Phase'. IEEE Transactions on Pattern Analysis and Machine Intelligence