Differential Privacy for Machine Learning

Anastasia Pustozerova
APustozerova@sba-research.org

SBA Research, Floragasse 7, 1040 Viennaa, Austria
&
Universität Wien, 1010 Vienna, Austria

Heutzutage spielt Machine Learning in verschiedenen Branchen eine zunehmend wichtigere Rolle. Im Gesundheitswesen beispielsweise werden große Datenmengen benötigt, um die Vorhersage von Krankheiten zu verbessern oder neue Medikamente zu erforschen. Die Verwendung dieser Daten ist oft mit Datenschutzbedenken verbunden, insbesondere wenn sie sensible Informationen über Einzelpersonen enthalten. Vorschriften wie die Datenschutzgrundverordnung (DSGVO) motivieren darüber hinaus die Forschung im Bereich Privacy-Preserving Machine Learning.

Differential Privacy (DP) wurde von Dwork et al. [1] vorgeschlagen, um Datensätze in einer Datenbank mit sensiblen Informationen zu schützen, während gleichzeitig Statistiken über die Daten abgefragt werden können. DP schützt die Privatsphäre des Einzelnen durch Hinzufügen von Rauschen zu den Abfrageergebnissen. Ziel ist es, dass sich ein Algorithmus bei Daten, die sich nur in einem Element unterscheiden, ähnlich verhält, d. h. ein ähnliches, kaum unterscheidbares Ergebnis liefert. Einer der Hauptvorteile von Differential Privacy ist die Möglichkeit, den Verlust der Privatsphäre zu quantifizieren. Im Jahr 2016 stellten Abadi et al. [2] eine differenziell private Version des Stochastic Gradient Descent (DP-SGD) vor, die nach wie vor eine der beliebtesten Methoden zum Schutz der Privatsphäre im Machine Learning ist. DP kann jedoch in verschiedenen Phasen eines maschinellen Lernprozesses erreicht werden, z. B. durch Hinzufügen von Rauschen zu den Trainingsdaten, zu den Gradienten (wie bei DP-SGD), zur Zielfunktion oder zum fertig trainierten Modell. Das Hinzufügen von Rauschen bei DP zur Verbesserung der Privatsphäre, führt jedoch unweigerlich zu einer Verringerung des Nutzens, weshalb bei der Anwendung von DP ein Kompromiss zwischen Privatsphäre und Nutzen eingegangen werden muss [3].

Differential Privacy ist auch in Szenarien relevant, in denen Daten zunächst dezentral gesammelt und gespeichert werden, aber kollaborativ analysiert werden sollen, z. B. beim föderierten Lernen (federated learning) [3]. Je nach Anforderung kann Differential Privacy in diesem Umfeld auf verschiedene Weise angewendet werden:

- **Central DP** bezieht sich auf eine Situation, in der ein vertrauenswürdiger Datenaggregator Zugang zu den Rohdaten hat und DP nur auf die Analyseergebnisse anwendet, z. B. das Machine Learning-Modell oder aggregierte Statistiken.
- **Local DP** wird vor der Aggregation angewendet, wenn die Nutzer dem Aggregator nicht vertrauen. Local DP bietet daher mehr Datenschutz als Central DP. Gleichzeitig führt Local DP zu einem stärkeren Rückgang des Nutzens der Daten und der Genauigkeit der Machine Learning-Modelle.
- **Distributed DP** zielt darauf ab, ähnliche Datenschutzgarantien wie bei Local DP zu bieten und gleichzeitig den Verlust an Nutzen auf das Niveau von Central DP zu reduzieren. Distributed DP kann z. B. durch die Verwendung sicherer

Aggregationsprotokolle wie Secure Multiparty Computation erreicht werden, verursacht aber auch zusätzliche Rechen- und Kommunikationskosten.

Die Verringerung des Nutzens ist eine der größten Herausforderungen der Differential Privacy. Wenn der Nutzen optimiert werden soll, muss besonders darauf geachtet werden, den Parameter für den Verlust der Privatsphäre so zu wählen, dass Differential Privacy die notwendigen Garantien für die Privatsphäre bietet. Es gibt keine gesetzlichen Richtlinien für die korrekte Anwendung von DP, dennoch wird DP bereits aktiv als Methode, z. B. von Apple, Google und Microsoft eingesetzt.

Meine Forschung konzentriert sich auf ein besseres Verständnis der Auswirkungen von Differential Privacy auf die Privatsphäre und den Nutzen von Machine Learning-Modellen. Insbesondere analysiere und vergleiche ich verschiedene Strategien zur Anwendung von DP im föderierten Lernen, um optimale Wege zur Nutzung von DP zu finden [3]. Die Erforschung datenschutzfreundlicher Techniken wie Differential Privacy ist für eine nachhaltige Entwicklung des Machine Learning unerlässlich, damit wir als Benutzer nicht mit unserer Privatsphäre für den Fortschritt bezahlen müssen.

[1] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., & Naor, M. (2006a). Our data, ourselves: Privacy via distributed noise generation. In Advances in Cryptology–EUROCRYPT.

[2] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L., 2016. Deep Learning with Differential Privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.

[3] Anastasia Pustozerova, Jan Baumbach, and Rudolf Mayer. Differentially Private Federated Learning: Privacy and Utility Analysis of Output Perturbation and DP-SGD. In *2023 IEEE International Conference on Big Data (Big Data)*.