Initialization Matters: Privacy-Utility Analysis of Overparameterized Neural Networks

Qiuyuan Ye (qiuyuan@comp.nus.edu.sg), Zhenyu Zhu (zhenyu.zhu@epfl.ch), Fanghui Liu (fanghui.liu@epfl.ch), Reza Shokri (reza@comp.nus.edu.sg) and Volkan Cevher (volkan.cevher@epfl.ch)

Overparameterization of Neural Network
Increasing Depth and Width of the Hidden layers (for DNN)

KL privacy bound for overparameterized DNN (informal)
Let $M_T$ be the per-record gradients subspace in Langevin diffusion with time $T$. Then constants $\beta, \sigma$ specify a relaxed smoothness condition. Then
\[
KL(W_{0:T} || W_{0:T}) = \frac{1}{2T} \int_0^T \left( \| \nabla L(W_t : D) - \nabla L(W_t : D^\prime) \|^2 \right) dt
\]

Initialization matters for small training time $T$

KL privacy bound for overparameterized DNN (informal)

\[
\frac{1}{2T} \int_0^T \left( \| \nabla L(W_t : D) - \nabla L(W_t : D^\prime) \|^2 \right) dt
\]

Implementation of Excess Empirical Risk

\[
\frac{1}{2T} \int_0^T \left( \| \nabla L(W_t : D) - \nabla L(W_t : D^\prime) \|^2 \right) dt
\]

**Main Takeaways**

- We theoretically prove and numerically show that for training DNNs with a small time, and for training linearized networks with any time
- Increasing width always hurts KL privacy
- Increasing depth helps KL privacy under certain initializations
- Under certain data regularity and large enough widths, we further prove privacy-utility trade-offs for training linearized networks and prove that it similarly relies on the choice of initialization distributions

**Special Case: Privacy-Utility Trade-offs for Training Linearized Network**

- Consider a linearized network by first-order Taylor expansion
  \[
  f_{W_{0:n}}(x) = f_{W_{0:n-1}}(x) + \frac{\partial f_{W_{0:n}}(x)}{\partial W_{0:n-1}} (W_{0:n} - W_{0:n-1})
  \]
- Under GD, DNN can work in the lazy training regime, under which this linearized network well approximates DNN training
- Theorem: For single output linearized network with hidden layer width $m$, bounded data with dimension $d$, under certain regularity conditions, if $d, m = \Omega(n)$ where $n$ is size of training dataset, then

**Numerical evidence for KL privacy loss of DNNs**

**Certainty:**

- All claims are proven, we list 20 epochs for each run, and then improve (decreases) with increasing depth $L$.

**Algorithmically:**

- The numerical KL privacy loss decreases under increasing depth $L$.

**Contribution:**

- KL privacy under simplified privacy model
- KL privacy under simplified privacy model

**Certainty:**

- All claims are proven, we list 20 epochs for each run, and then improve (decreases) with increasing depth $L$. **Algorithmically:** The numerical KL privacy loss decreases under increasing depth $L$.

**Acknowledgements:** This work was supported by Hasler Foundation Program: Hasler Responsible AI (project number 211043), the Swiss National Science Foundation (SNSF) under grant number 200011_205011, Google PDFO faculty research award, Intel within the www.private-ai.org center, Meta faculty research award, the NUS Early Career Research Award (NUS ECRCA award number NUS ECRCA FY19 P16), and the National Research Foundation, Singapore under its Strategic Capability Research Centres Funding Initiative.