Challenger 2: Certified Data Removal from Machine Learning Models

Weijie Guan

Weakness

1. Limited application scenarios

1. The theory in the paper is based on **linear models and convex loss functions**, which limits the application scenarios, for example, the paper only use logistic-regression-based model to verify the effectiveness even on multiclassification problem.

2. The paper mentions removing data from the machine learning model, but it seems like Certified Removal only remove data **on a simple linear downstream model**, and **does not remove the data at the source**, for example, there is **a potential risk of data leakage in the upstream encoder** that cannot be addressed or handled by this approach

2. Incomplete experiments

1. The experiment did not independently explore the effects of λ and σ.
    1. **How do the value of λ and σ affect the expected number of supported removals independently?**
    2. **How do the value of λ and σ affect the effectiveness of Certified Removal?**

2. The experiment lacks proof of the effectiveness of **loss perturbation** and direct verification of the effect of **Certified Removal**, for example, comparing the prediction uncertainty for the removed samples.
    1. **How to evaluate that whether data points are actually effectively removed from the model?**

Weakness

3. Complexity

1. It seems like Certified Removal needs to calculate the Hessian matrix and its inversion which leads to high complexity. **Is it possible that retraining the model is a faster way than Certified Removal especially when a lot of data needs to be removed?**