Robust Statistics

March 6, 2022

Contents

1 Part 1

- Can we develop learning algorithms that are robust to a constant fraction of corruptions in the data
- Statistical Learning Problem: Input: sample generated by a **statistical model** with unknown θ^*. Goal is to estimate parameters θ such that $\theta \approx \theta^*$
- Strong contamination model: Let $\mathcal{F}$ be a family of statistical models. We say that a set of N samples is ϵ-corrupted from $\mathcal{F}$ if it is generated as follows:
  - N samples drawn from unknown $F \in \mathcal{F}$
  - omniscient adverdary inspects samples and arbitrarily changees an ϵ-fraction of them
- Example: Parameter estimation
  - Given i.i.d samples from unknown distribution, how do we estimate its parameters?
  - mean: $\frac{1}{N}\sum_{i=1}^{N} X_i \to \mu$, empirical variance: $\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 \to \sigma^2$
  -

References

[1] Tutorial: Recent Advances in High-Dimensional Robust Statistics, ICML 2020 `http://www.iliasdiakonikolas.org/icml-robust-tutorial.html`