List-decodable mean estimation: the problem

Input: Dataset \( \mathcal{X} \) of size \( n \): a fraction i.i.d. from bounded covariance \( D \), others arbitrary.

Output: hypothesis list \( L \) of size \( O(k) := O(\sqrt{n}) \) containing at least one “good” mean estimate.

Lower bound: list size \( O(k) \), \( O(dk) \) samples, and estimation error \( O(\sqrt{n}/\log n) \) necessary.

Flexible, robust statistical model!
- Robust analog of learning mixture models
- Apps: crowd-sourcing, community detection
- Semi-verified learning: “a data prism”

Our approaches: simpler and faster list-decodable mean estimation

I. Robustly matching the “PCA barrier”: [DKKLT20]

Decoupling from subspace identification.

Potential function: \( O(\text{th}) \) largest eigenvalue. Identifiability proof + naive sampling gets us the rest of the way.

A new approach: “decoupling” clustering and filtering


Fast filtering via Ky Fan SDP regret minimization.

II. Robustly breaking the “PCA barrier”: [DKKLT21]

A new, “one-shot” MMW potential alternative.

Rethinking the [DKK20] multifilter.

New multifilter tree (one layer)

Putting it all together.

Previous approaches

“Filtering”: [DKKMS17, Sei18]

Input: Dataset \( \mathcal{X} \) of size \( n \): a fraction i.i.d. from bounded covariance \( D \), others arbitrary.

Output: hypothesis list \( L \) of size \( O(k) := O(\sqrt{n}) \) containing at least one “good” mean estimate.

Lower bound: list size \( O(k) \), \( O(dk) \) samples, and estimation error \( O(\sqrt{n}/\log n) \) necessary.

Flexible, robust statistical model!
- Robust analog of learning mixture models
- Apps: crowd-sourcing, community detection
- Semi-verified learning: “a data prism”

I. Robustly matching the “PCA barrier”: [DKKLT20]

Decoupling from subspace identification.

Potential function: \( O(\text{th}) \) largest eigenvalue. Identifiability proof + naive sampling gets us the rest of the way.

A new approach: “decoupling” clustering and filtering


Fast filtering via Ky Fan SDP regret minimization.

II. Robustly breaking the “PCA barrier”: [DKKLT21]

A new, “one-shot” MMW potential alternative.

Rethinking the [DKK20] multifilter.

New multifilter tree (one layer)

Putting it all together.