Mixed Regression: Minimax Optimal Rates

Constantine Caramanis

The University of Texas at Austin
constantine@utexas.edu

Joint work with Yudong Chen and Xinyang Yi

June 13, 2014
a simple problem

\[ y_i = \langle x_i, \beta^* \rangle + e_i, \quad i = 1, \ldots, n, \]

- \( \beta^* \in \mathbb{R}^p \)
- statistics: \( n \geq p \), error \( \sim \sigma \sqrt{p/n} \).
- computation: min : \( \| y - X \beta \|_2 \).
sparse version: $\beta^* \in \mathbb{R}^p$, sparse.

low-rank version: $\beta^* \in \mathbb{R}^{p \times p}$, low-rank.

low-rank plus sparse: $\beta^* \in \mathbb{R}^{p \times p}$, $\beta^* = L + S$.

low-rank plus sparse plus column sparse: $\beta^* \in \mathbb{R}^{p \times p}$, $\beta^* = L + S + C$.

e.tc.

mixture: $\beta^* = \beta_1^*$ or $\beta^* = \beta_2^*$ ?
a mixture problem

\[ y_i = z_i \cdot \langle x_i, \beta_1^* \rangle + (1 - z_i) \cdot \langle x_i, \beta_2^* \rangle + e_i, \quad i = 1, \ldots, n, \]

\( \beta_1^*, \beta_2^* \in \mathbb{R}^p, \quad z_i \in \{0, 1\}. \)
if we don’t care about computational complexity, (often) it’s easy.

if we don’t care about sample complexity, (sometimes) it’s easy.

if we care about both...
- exact solution seems to be hard \((\text{Subset-Sum})\).

- classical: expectation maximization – guess labels, find \((\beta_1^*, \beta_2^*)\), repeat.

- tensor approach.
our results: optimal rates

\[ y_i = z_i \cdot \langle x_i, \beta_1^* \rangle + (1 - z_i) \cdot \langle x_i, \beta_2^* \rangle + e_i, \quad i = 1, \ldots, n. \]

a convex formulation such that: if \( x_i \) independent, sub-Gaussian,

- minimax-optimal rates when \( \{e_i\} \) arbitrary norm-bounded.
- minimax-optimal rates when \( \{e_i\} \) sub-Gaussian, and balanced mixture.
a convex formulation

\[ K^* = \frac{1}{2}(\beta_1^* \beta_2^* + \beta_2^* \beta_1^*) \]

\[ g^* = \frac{1}{2}(\beta_1^* + \beta_2^*). \]

given \((K^*, g^*)\),

\[ J^* = g^* g^* \top - K^* = \frac{1}{4}(\beta_1^* - \beta_2^*)(\beta_1^* - \beta_2^*) \top. \]
key features of the results

- arbitrary noise: minimax error rate:
  \[
  \frac{\|\text{noise}\|}{\sqrt{n}}.
  \]

- stochastic noise: minimax rate changes in high/low SNR regimes.
  interpolates between:
  \[
  \sigma \left( \frac{p}{n} \right)^{1/2} \quad \longleftrightarrow \quad \sigma \left( \frac{p}{n} \right)^{1/4}.
  \]

  high SNR

  low SNR