Low-rank Interaction with Sparse Additive Effects
Model for Large Data Frames

EXISTING TOOLS: Low-rank models [3, 4, 5]

- Rows/columns embedded in low-dimensional space
- Latent factors
- Summary features/individuals

$$M_0^{ij} = \sum_{k=1}^{K} U_k V_{jk}$$

Multiplicative Interactions

MISSING EXTENSION: main effects (individuals, groups, covariates)

$$M_0^{ij} = \sum_{k=1}^{q} c_k X(k)_{ij} + \Theta_{ij}$$

- Sparse regression on a dictionary of covariates
- Low-rank design

Statistically + computationally efficient way to estimate
Main effects and interactions in mixed data frames

ESTIMATION: Sparse Additive Effects and Low-rank Interactions

$$[\alpha, \Theta] \in \arg \min_{\alpha, \Theta} \mathcal{L}(M) + \sum_{k=1}^{d} \lambda_k \|\alpha_k\|_1 + \lambda_c \|\Theta\|_F.$$ 

- Assumptions:
- Dictionary matrices: \( \sum_k |X(k)| \leq \infty \)
- Link functions strongly convex
- Subexponential observations
- Minimum probability of observing an entry \( \pi > 0 \)

NUMERICAL RESULTS: Timing and estimation

Theorem: the MCGD method converges to an \( \epsilon \)-solution in \( O(1/\epsilon) \) iterations.

OPTIMIZATION: Mixed coordinate gradient descent

- Augmented problem: \( \min_{(\Theta, \alpha)} \mathcal{L}(\alpha, \Theta) + \lambda_\alpha \|\alpha\|_1 + \lambda_c R \) s.t. \( R \geq R \geq |\Theta| \).

Algorithm 1 MCGD Method:

1. Initialize: \( \{\Theta(0), \alpha(0), R(0)\} \)
2. For \( t = 1, 2, \ldots, T \) do
3. // Update for \( \alpha // \)
   \[ \alpha^{(t)} = T_{\pi} \left( \alpha^{(t-1)} - \frac{\nabla \mathcal{L}(\alpha^{(t-1)}; \Theta^{(t-1)})}{\lambda} \right) \]
   where \( T_{\pi}() \) is the soft-threshold operator.
4. // Update for \( \Theta; R // \)
   Update the upper bound as: \( R^{(t)} = \lambda R^{(t-1)} \).
5. Obtain the update direction \( (\Theta^{(t)}, R^{(t)}) \) as \( (\Theta^{(t)}, R^{(t)}) = \left\{ \begin{align*}
\Theta^{(t)}, R^{(t)} & , \text{if } \lambda c \geq |\nabla \mathcal{L}(\Theta^{(t)}, \alpha^{(t)})|, \\
\Theta^{(t)} & , \text{if } \lambda c < |\nabla \mathcal{L}(\Theta^{(t)}, \alpha^{(t)})|.
\end{align*} \right. \)
6. Obtain the CG update as \( (\Theta^{(t)}, R^{(t)}) = (\Theta^{(t)}, R^{(t-1)}) + \beta(\Theta^{(t)}, R^{(t)} - \Theta^{(t-1)}, R^{(t-1)} - R^{(t-1)}) \)

7. end for
8. Return: \( (\Theta^{(T)}, R^{(T)}) \).

Theorems:

- Total per iteration cost: \( O(\|\mathbf{E}\| + \|\mathbf{R}(\max(n, p) \log(1/\epsilon) + g)) \).
- Provably faster than the sublinear rate in [6].

The authors would like to thank for the useful comments from three anonymous reviewers.

Acknowledgement

The work was supported by the grant NSF CCF-BSF 1714672.

References:


