NIPS 2011

On Learning Discrete Graphical Models Using Greedy Methods

Ali Jalali, Christopher C. Johnson and Pradeep Ravikumar

University of Texas at Austin

- Two Contributions ::
  - Statistical estimation with **sparse** parameters :: analysis of forward-backward **greedy algorithm**; better than ell_1!
  - Application to Discrete Graphical Model Selection

Statistical Model :: $X^{(i)} \sim P(X;\theta^*)$

Neg. Log-Likelihood :: $\mathcal{L}(\theta;D) = \frac{1}{n}\sum_{i=1}^{n} -\log P(X^{(i)};\theta)$

Sparsity Constrained MLE :: $\min_{\theta:\|\theta\|_0 \leq k} \mathcal{L}(\theta;D)$

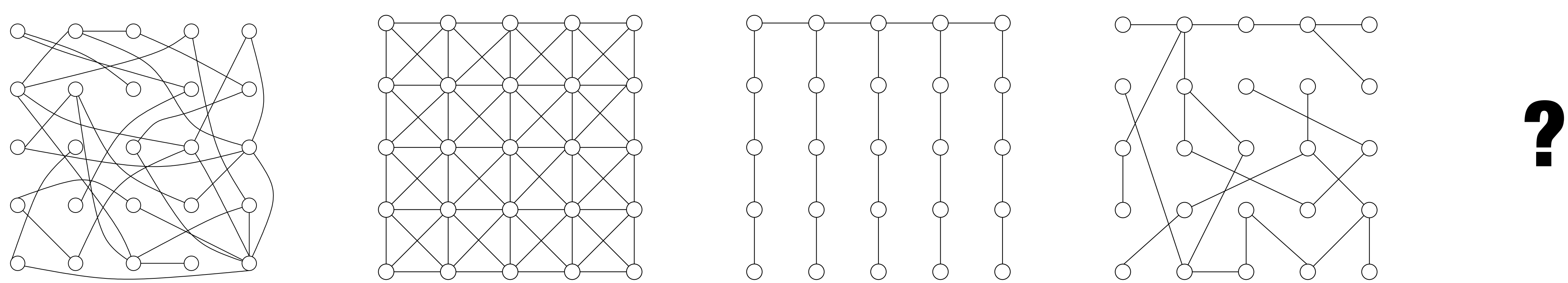
ell_1?

Bias; Restrictive Model Conditions

Non-Convex?

Learning Discrete Graphical Models

- Discrete Random Variables $X = (X_1, X_2, \ldots, X_p)$
- Discrete Graphical Model :: $P(X; \theta, \mathbf{G}) \propto \exp\left\{ \sum_{(s,t) \in E(\mathbf{G})} \theta_{st} \phi_{st}(x_s, x_t) \right\}$
- Given :: **n** samples $D := (X^{(1)}, \ldots, X^{(n)})$ where $X^{(i)} \sim P(X; \theta^*, \mathbf{G})$
- Problem :: Estimate underlying graph **G**



Forward-Backward Greedy Algorithm

- Generalization of [T. Zhang, 2008] greedy algorithm for linear regression to general sparse statistical estimation

- Algorithm (Stopping Threshold \epsilon) ::

    - Forward: Find best co-ordinate to add ; add if improvement greater than \epsilon; set \delta = amount of improvement

    - Backward: Prune co-ordinates with loss-increase smaller than \delta

Theorem [Sparsistency]: Recovers support of true parameter, given restricted strong convexity, sufficient stopping threshold

Comparison: Learning Discrete Graphical Models

	ell_1	**greedy**
Model Assumptions	Irrepresentable / Incoherence	Restricted Strong Convexity
Sample Complexity	$n = \Omega(d^3 \log p)$	$n = \Omega(d^2 \log p)$
Comp. Complexity	$O(p^4)$	$O(d^3\, p^2)$

Better in all respects!

i.e. don't use ell_1 regularization; use greedy!

Oh, and
Information-theoretically Optimal
(Santhanam, Wainwright 08)

Visit us at Poster