Joint Gaussian Graphical Model Series – VIII

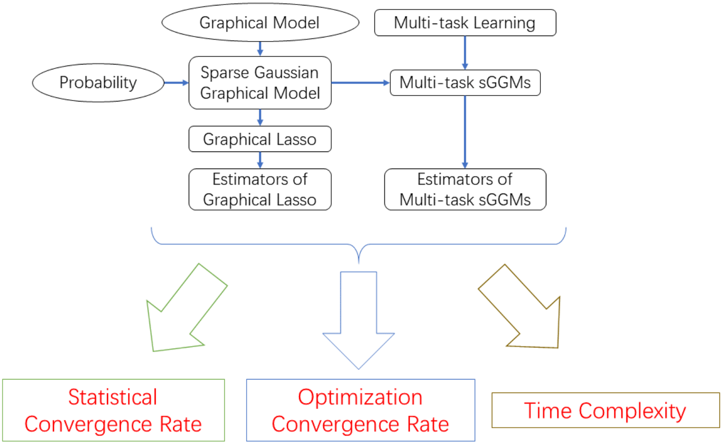
A deep introduction of the metrics for evaluating an/a estimator/learner

Beilun Wang
Advisor: Yanjun Qi

[1]Department of Computer Science, University of Virginia
http://jointggm.org/

Sep 22nd, 2017

Road Map


Graphical Model
Multi-task Learning
Probability
Sparse Gaussian Graphical Model
Multi-task sGGMs
Graphical Lasso
Estimators of Graphical Lasso
Estimators of Multi-task sGGMs
Statistical Convergence Rate
Optimization Convergence Rate
Time Complexity

Outline

Notation

Notation

- X The data matrix
- Σ The covariance matrix.
- Ω The precision matrix.
- p The number of features.
- n The number of samples in the data matrix.
- s The number of non-zero entries in the precision matrix.

Review

Review from last talk

- We introduce different sGGM estimators and their solution.

Review from last talk

- We introduce different sGGM estimators and their solution.
- We briefly introduce the three metrics used in evaluating an estimator.

Review from last talk

- We introduce different sGGM estimators and their solution.
- We briefly introduce the three metrics used in evaluating an estimator.
- We introduce different multi-task sGGMs estimators and their optimization challenges.

The metrics for evaluating an estimator

Motivation I: Select a proper estimator

- There may be a lot of similar estimators.

Motivation I: Select a proper estimator

- There may be a lot of similar estimators.
- You need to decide which one to use.

Motivation I: Select a proper estimator

- There may be a lot of similar estimators.
- You need to decide which one to use.
- You need some metrics to make the decision.

Motivation II: Evaluate a novel method

- You may come out a new estimator.

Motivation II: Evaluate a novel method

- You may come out a new estimator.
- You want to know whether this novel estimator is no worse than the previous ones.

Motivation II: Evaluate a novel method

- You may come out a new estimator.
- You want to know whether this novel estimator is no worse than the previous ones.
- Then you need some metrics to evaluate the estimator.

Background: Two major properties

- Two major properties: Accuracy and Speed.

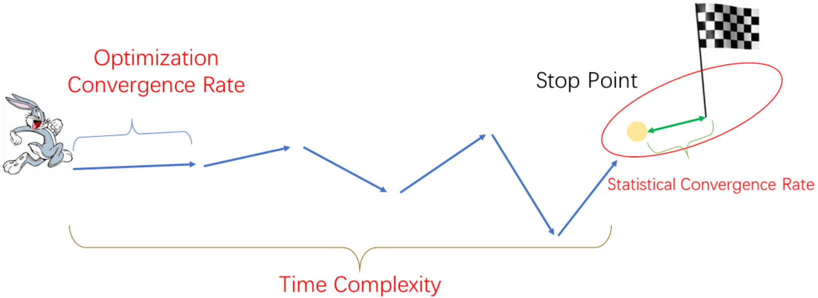
Background: Two major properties

- Two major properties: Accuracy and Speed.
- Accuracy:
  - Statistical Convergence rate
  - how close to the Truth
  - Statisticians

Background: Two major properties

- Two major properties: Accuracy and Speed.
- Accuracy:
  - Statistical Convergence rate
  - how close to the Truth
  - Statisticians
- Speed:
  - Optimization convergence rate
  - Optimization researchers
  - Computational complexity
  - Computer Scientists

Overview Figure



Statistical Convergence Rate

Statistical Convergence Rate : Definition

- The task for an estimator is parameter estimation.

Statistical Convergence Rate : Definition

- The task for an estimator is parameter estimation.
- Suppose the parameter you need to estimate is θ, the truth is θ^*

Statistical Convergence Rate : Definition

- The task for an estimator is parameter estimation.
- Suppose the parameter you need to estimate is θ, the truth is θ^*
- $\| \theta - \theta^* \|$ or $\mathcal{R}(\theta - \theta^*)$

A simple example: Estimate the mean

On the whiteboard.

Elementary Estimator[Yang et al.(2014b)Yang, Lozano, and Ravikumar]

$$\underset{\theta}{\text{argmin}} \mathcal{R}(\theta) \tag{4.1}$$

$$\text{Subject to: } \mathcal{R}^*(\theta - \mathcal{B}^*(\widehat{\phi})) \leq \lambda_n \tag{4.2}$$

Here $\mathcal{B}^*(\widehat{\phi})$ is a backward mapping for $\widehat{\phi}$.

Example: sparse linear regression[Yang et al.(2014a)Yang, Lozano, and Ravikumar]

$$\underset{\theta}{\operatorname{argmin}} ||\theta||_1 \tag{4.3}$$

$$\text{Subject to: } ||\theta - (X^T X + \epsilon I)^{-1} X^T y||_\infty \leq \lambda_n \tag{4.4}$$

Hands on: Elementary Estimator for high-dimensional linear regression

On the whiteboard.

Hands on: DIFFEE

On the whiteboard.

Conclusions

- In high-dimensional setting, related to $\frac{\log p}{n}$.
- Equivalent estimators still have differences in constants or constraints

Optimization Convergence Rate

Optimization Convergence Rate : Definition

- Linearly Converge: $\lim_{k \to \infty} \frac{|\theta_{k+1} - L|}{|\theta_k - L|} = \mu_k$

Optimization Convergence Rate : Definition

- Linearly Converge: $\lim_{k \to \infty} \frac{|\theta_{k+1} - L|}{|\theta_k - L|} = \mu_k$
-
  - Linearly, if $\mu_k \in (0, 1)$
  - Superlinearly, if $\mu_k \to 0$ when $k \to \infty$.
  - Sublinearly, if $\mu_k \to 1$ when $k \to \infty$

Optimization Convergence Rate : Definition

- Linearly Converge: $\lim_{k \to \infty} \frac{|\theta_{k+1} - L|}{|\theta_k - L|} = \mu_k$
-
  - Linearly, if $\mu_k \in (0, 1)$
  - Superlinearly, if $\mu_k \to 0$ when $k \to \infty$.
  - Sublinearly, if $\mu_k \to 1$ when $k \to \infty$
- Higher order: $\lim_{k \to \infty} \frac{|x_{k+1} - L|}{|x_k - L|^q} > 0$.

Optimization Convergence Rate : Definition

- Linearly Converge: $\lim_{k \to \infty} \frac{|\theta_{k+1} - L|}{|\theta_k - L|} = \mu_k$
-
  - Linearly, if $\mu_k \in (0, 1)$
  - Superlinearly, if $\mu_k \to 0$ when $k \to \infty$.
  - Sublinearly, if $\mu_k \to 1$ when $k \to \infty$
- Higher order: $\lim_{k \to \infty} \frac{|x_{k+1} - L|}{|x_k - L|^q} > 0$.
- Closed form solution

Optimization Convergence Rate : Definition

- Linearly Converge: $\lim_{k \to \infty} \frac{|\theta_{k+1} - L|}{|\theta_k - L|} = \mu_k$
-
  - Linearly, if $\mu_k \in (0, 1)$
  - Superlinearly, if $\mu_k \to 0$ when $k \to \infty$.
  - Sublinearly, if $\mu_k \to 1$ when $k \to \infty$
- Higher order: $\lim_{k \to \infty} \frac{|x_{k+1} - L|}{|x_k - L|^q} > 0$.
- Closed form solution
- Closed form $\geq$ Higher order $\geq$ linear

Optimization Convergence Rate: Basic Results

- Gradient Descent based method: Linear

Optimization Convergence Rate: Basic Results

- Gradient Descent based method: Linear
-
  - gradient descent
  - SGD
  - ADMM / proximal gradient descent

Optimization Convergence Rate: Basic Results

- Gradient Descent based method: Linear
-
  - gradient descent
  - SGD
  - ADMM / proximal gradient descent
- Newton method based method: Quadratic

Optimization Convergence Rate: Basic Results

- Gradient Descent based method: Linear
-
  - gradient descent
  - SGD
  - ADMM / proximal gradient descent
- Newton method based method: Quadratic
- Elementary Estimator: Closed form solution

Optimization Convergence Rate: Different methods

	Single sGGM			Multiple sGGMs	
Method:	GLasso	CLIME	EEGM	JGL	FASJEM
Rate of Convergence	Linear	NA	Closed form	Linear	Linear

Computational Complexity

Computational Complexity: Definition

- Complexity of an algorithm is the amount of resources required for running it.

Computational Complexity: Definition

- Complexity of an algorithm is the amount of resources required for running it.
- In machine learning, it is mainly related to n and p.

Computational Complexity: Definition

- Complexity of an algorithm is the amount of resources required for running it.
- In machine learning, it is mainly related to n and p.
- Use big O notation

Computational Complexity: how to calculate

- Some cases:
  - Matrix Multiplication: $O(np^2)$
  - Matrix inversion $O(p^3)$
  - SVD inversion $O(p^3)$
  - soft-thresholding $O(p^2)$

Computational Complexity: how to calculate

- Some cases:
  - Matrix Multiplication: $O(np^2)$
  - Matrix inversion $O(p^3)$
  - SVD inversion $O(p^3)$
  - soft-thresholding $O(p^2)$
- How to calculate:
  - Num of Iter $\times$ Computational complexity of each Iter
  - Direct calculate e.g., Closed form solution
  - Use existing method e.g., linear programming
  - Special case: linear convergence.

Computational Complexity: Different methods

	Single sGGM			Multiple sGGMs		
Method:	GLasso	CLIME	EEGM	JGL	FASJEM	SIMUL
Computational Complexity	$O(Tp^2)$	$O(p^5)$	$O(p^2)$	$O(Tp^3)$	$O(Tp^2)$	$O(K^4 p$

Summary

- We introduce the statistical convergence rate.
- We introduce the optimization convergence rate.
- We introduce the computational complexity.

References I

E. Yang, A. Lozano, and P. Ravikumar.
Elementary estimators for high-dimensional linear regression.
In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 388–396, 2014a.

E. Yang, A. C. Lozano, and P. K. Ravikumar.
Elementary estimators for graphical models.
In *Advances in Neural Information Processing Systems*, pages 2159–2167, 2014b.