Корреляционная структура

Многомерные гауссовы модели. Обратная к ковариационной матрица. Разреживание. Блочный оптимизационный подход. Графическое Lasso.

ПМИ ФКН ВШЭ, 1 декабря 2018 г.

Алексей Артемов[1,2]

[1] Сколтех [2] ФКН ВШЭ

Оглавление

- Многомерные гауссовы системы
- Факторизация совместных распределений наблюдаемых
- Понятие графических моделей
- Выборочная ковариационная матрица и ℓ_1-регуляризация оценки
- Блочно-координатный метод оптимизации регуляризованного правдоподобия
- Lasso-метод оптимизации регуляризованного правдоподобия
- Алгоритм QUIC

Мотивация выявления структуры

Мотивация выявления структуры

- Размерность задач в современной статистике чрезвычайно высокая
  - 10^3 котировок акций, 10^6 вокселей МРТ, 10^8 пользователей Facebook, ...
  - Объем выборки даже бывает меньше размерности!
- Естественно желание снизить размерность описания соответствующих задач
- Один из способов это сделать — исключить корреляционные зависимости между некоторыми наблюдаемыми

Мотивация выявления структуры

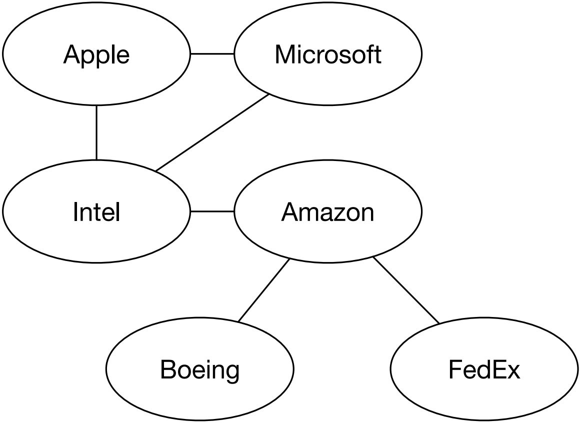


Рис.: Пример зависимости между 6 переменными (наличие связи обозначается ребром в графе).
Hallac D. et al. Network inference via the time-varying graphical lasso //Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – ACM, 2017. – С. 205-213. MLA

Мотивация выявления структуры

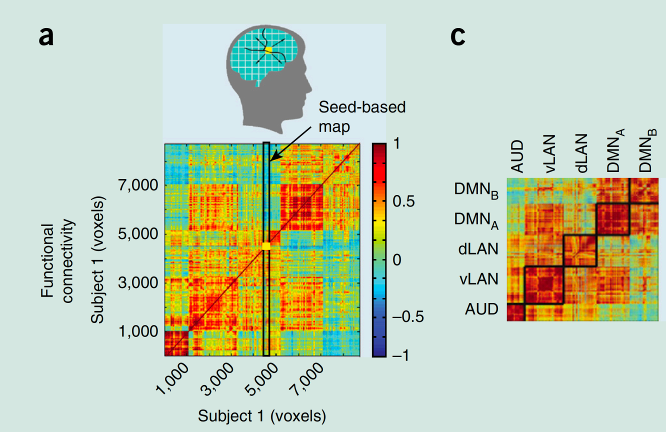


Рис.: Функциональная связность при полном корреляционном повоксельном анализе: **a:** корреляция между выбранным RoI (желтый цвет) и остальными вокселями; **c:** повоксельные ковариационные матрицы описывают организацию подсетей у испытуемого.

Cohen, Jonathan D., et al. "Computational approaches to fMRI analysis." Nature neuroscience 20.3 (2017): 304.

Многомерные гауссовы системы

Многомерные гауссовы системы

Определение (Многомерный гауссовский вектор)

Это случайный вектор $\boldsymbol{x} \in \mathbb{R}^n$, имеющий распределение вида

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

- **Пример:** двумерное нормальное распределение ($\boldsymbol{x} = (x_1, x_2)^\intercal \in \mathbb{R}^2$) имеет плотность

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \times \right.$$

$$\left. \times \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \rho \frac{2(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right] \right\}$$

Многомерные гауссовы системы

- Соответствующая ковариационная матрица равна

$$\mathbf{\Sigma} = \mathrm{E}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\mathsf{T}] = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

- Случай $\rho = 0$: отсутствие корреляций (независимость!)
- Как запишется плотность $\boldsymbol{x} \in \mathbb{R}^2$?

Многомерные гауссовы системы

- Соответствующая ковариационная матрица равна

$$\boldsymbol{\Sigma} = \mathrm{E}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^\intercal] = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

- Случай $\rho = 0$: отсутствие корреляций (независимость!)
- Как запишется плотность $\boldsymbol{x} \in \mathbb{R}^2$?
- Ковариационная матрица диагонализуется: $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$, а для плотности выполнена **факторизация**

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{ -\frac{1}{2}\left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}\right] \right\} =$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{ -\frac{(x_1 - \mu_1)^2}{2\sigma_1^2} \right\}$$

$$\times \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{ -\frac{(x_2 - \mu_2)^2}{2\sigma_2^2} \right\}$$

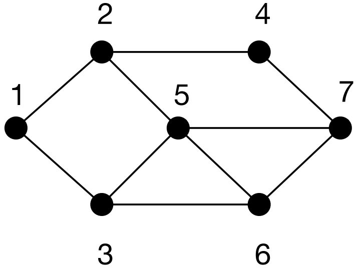
Понятие графической модели

- Говорят, что распределение P факторизуется по ненаправленному графу G, если для совместной плотности $f(\boldsymbol{x})$ выполнено:

$$f(\boldsymbol{x}) = Z^{-1} \prod_{A \in \mathcal{A}} \phi_A(\boldsymbol{x}_A),$$

  где $\mathcal{A}$ – подмножества графа
- Здесь переменные $x = \{x_v, v \in V\}$, $\boldsymbol{x}_A = \{x_v, v \in A\}$ индексируются узлами графа
- Такая факторизация соответствует марковскому свойству: $A \perp\!\!\!\perp B | S$, если S отделяет A от B в графе

Понятие графической модели



- Граф выше соответствует факторизации

$$f(\boldsymbol{x}) = \psi_{12}(x_1, x_2)\psi_{13}(x_1, x_3)\psi_{24}(x_2, x_4)\psi_{25}(x_2, x_5)\times$$
$$\times\ \psi_{356}(x_3, x_5, x_6)\psi_{47}(x_4, x_7)\psi_{567}(x_5, x_6, x_7)$$

Гауссовы графические модели

- Пусть $\boldsymbol{\mu} = 0$, тогда

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \exp\Big\{ -\frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{\Sigma}^{-1} \boldsymbol{x} \Big\}$$

$$\propto \det \boldsymbol{\Theta} \exp\Big\{ -\frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{\Theta} \boldsymbol{x} \Big\},$$

  где $\boldsymbol{\Sigma} = \mathrm{E}\left[\boldsymbol{x}\boldsymbol{x}^\intercal\right]$ и $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$
- Структура фиксируется в ковариационной матрице: если матричный элемент $\Theta_{ij} = 0$, то $x_i \perp\!\!\!\perp x_j$
- Условная независимость $\equiv$ разреженность

Задача оценивания корреляционной структуры

Стандартные статистические оценки

- Пусть задана выборка $\mathbf{X}^\ell = \{\boldsymbol{x}_i\}_{i=1}^\ell$
- Почему бы не использовать выборочную ковариационную матрицу:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{\ell - 1} \sum_{i=1}^{\ell} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^\intercal,$$

  где $\widehat{\boldsymbol{\mu}} = \frac{1}{\ell} \sum_{i=1}^{\ell} \boldsymbol{x}_i$ – выборочное среднее.
- Для гауссовских систем оценка $\widehat{\boldsymbol{\Sigma}}$ является несмещенной и состоятельной

Недостатки стандартных оценок

- Качество оценивания зависит от соотношения объема выборки ℓ и размерности пространства n:

$$\begin{cases} \text{если } \ell \gg n, & \text{то все хорошо,} \\ \text{если } \ell \leqslant n, & \text{то оценка } \widehat{\mathbf{\Sigma}} \text{ нестабильна.} \end{cases}$$

- Для $\ell \leqslant n$ использование выборочной оценки приводит к тому, что матрица $\widehat{\mathbf{\Sigma}}$ вырождена (невозможно вычислить обратную $\widehat{\mathbf{\Theta}} = \widehat{\mathbf{\Sigma}}^{-1}$) **(докажите!)**
- Стандартную оценку не получится использовать в разреженной модели (вообще говоря, все коэффициенты матрицы будут ненулевыми)
- Больше на `https://en.wikipedia.org/wiki/Estimation_of_covariance_matrices`

Неформальная постановка задачи

- Разумно иметь минимальную по сложности модель (например, графическую), которая адекватно описывает данные
- **Задача:** по заданной выборке данных $\mathbf{X}^\ell$ найти такую корреляционную матрицу, чтобы получить **как можно более разреженный** граф зависимостей между переменными
- Целью является как выявление (разреженной) матрицы зависимостей между переменными, так и получение регуляризованной оценки ковариационной матрицы
- Условие разреженность обеспечивается ℓ_1-регуляризацией решения

Формальная постановка задачи

- Пусть задана выборка $\mathbf{X}^\ell = \{\boldsymbol{x}_i\}_{i=1}^\ell, \boldsymbol{x}_i \in \mathbb{R}^n$. Запишем правдоподобие в предположениях гауссовской модели:

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{\ell n}{2}} \prod_{i=1}^{\ell} [\det \boldsymbol{\Sigma}]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu})^\mathsf{T} \Sigma^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})\right\}$$

  Эквивалентно (trace trick, **докажите!**):

$$\log \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{const} - \frac{\ell}{2} \log\det \boldsymbol{\Sigma} - \frac{1}{2} \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1} \boldsymbol{S}\right],$$

  где $\boldsymbol{S} = \sum_{i=1}^\ell (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})^\mathsf{T}$
- Стандартная задача вида $\log \mathcal{L} \to \max\limits_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ дает **(докажите!)**

$$\widehat{\boldsymbol{\mu}}_{\mathsf{MLE}} = \frac{1}{\ell} \sum_{i=1}^{\ell} \boldsymbol{x}_i, \qquad \widehat{\boldsymbol{\Sigma}}_{\mathsf{MLE}} = \frac{1}{\ell - 1} \sum_{i=1}^{\ell} (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\mathsf{MLE}})(\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}_{\mathsf{MLE}})^\mathsf{T},$$

Формальная постановка задачи

- Итак, стандартно $\log \mathcal{L} \to \max\limits_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ и решается задача на максимум:

$$\widehat{\boldsymbol{\Sigma}}_{\text{MLE}} = \underset{\boldsymbol{X} \text{ is PSD}}{\arg\max} \Big[\text{const} - \frac{\ell}{2} \log\det \boldsymbol{X} - \frac{1}{2} \operatorname{tr}(\boldsymbol{S}\boldsymbol{X}^{-}1) \Big],$$

- Для разреживания обратной ковариационной матрицы часто применяется ℓ_1-регуляризация
- Таким образом, требуется найти ℓ_1-регуляризованную ОМП:

$$\widehat{\boldsymbol{\Sigma}}_{\ell_1}^{-1} = \underset{\boldsymbol{X} \text{ is PSD}}{\arg\max} \Big[\log\det \boldsymbol{X} - \operatorname{tr}(\boldsymbol{S}\boldsymbol{X}) - \lambda \|\boldsymbol{X}\|_1 \Big]$$

- λ: баланс между максимизацией правдоподобия и разреженностью

Вспомним: ℓ_1-регуляризация

- L2 регуляризация для многомерной линейной регрессии

$$\|\boldsymbol{y} - \boldsymbol{X}\mathbf{w}\|^2 + \alpha\|\mathbf{w}\|^2 \to \min_{\mathbf{w}\in\mathbb{R}^d}$$

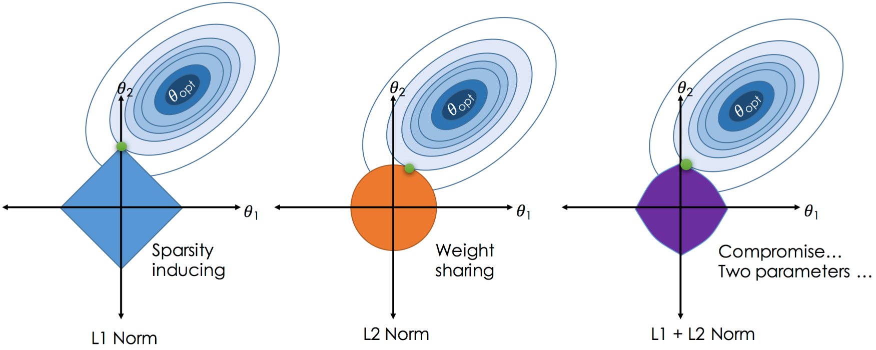
- L1 регуляризация (LASSO)

$$\|\boldsymbol{y} - \boldsymbol{X}\mathbf{w}\|^2 + \alpha\|\mathbf{w}\|_1 \to \min_{\mathbf{w}\in\mathbb{R}^d}$$

- L1/L2 регуляризация (Elastic Net)

$$\|\boldsymbol{y} - \boldsymbol{X}\mathbf{w}\|^2 + \alpha_1\|\mathbf{w}\|_1 + \alpha_2\|\mathbf{w}\|_2^2 \to \min_{\mathbf{w}\in\mathbb{R}^d}$$

Геометрическая интерпретация регуляризаторов



Изображение: http://www.ds100.org/sp17/assets/notebooks/linear_regression/Regularization.html

Интерпретация регуляризаторов

Рис.: Многомерное пространство параметров

Рис.: Регуляризованная модель

Известные подходы к получению решения

- Требуется найти ℓ_1-регуляризованную ОМП:

$$\widehat{\boldsymbol{\Sigma}}^{-1} = \underset{\boldsymbol{X} \text{ is PSD}}{\arg\max} \Big[\log\det \boldsymbol{X} - \mathrm{tr}(\boldsymbol{S}\boldsymbol{X}) - \lambda \|\boldsymbol{X}\|_1 \Big]$$

  - Разреженность, обратимость, …
- Блочно-координатный оптимизационный алгоритм [Banerjee et al., 2008]
- Графическое LASSO [Friedman et al., 2008]
- QUIC [Hsieh et al., 2014]

Оптимизация в пространстве матриц

Полезное определение

- Пусть $||\cdot||$ — норма в $\mathbb{R}^n$ и $\boldsymbol{z}, \boldsymbol{x} \in \mathbb{R}^n$.
- Сопряженная $||\cdot||$ «дуальная норма», обозначаемая $\|\cdot\|_*$, определяется как

$$||\boldsymbol{z}||_* = \sup_{||\boldsymbol{x}|| \leqslant 1} \boldsymbol{z}^\intercal \boldsymbol{x}$$

- Аналогичное определение существует для матричных норм

Алгоритм [Banerjee et al., 2008] (трюки)

- Требуется найти

$$\max_{\boldsymbol{X} \text{ is PSD}} \Big[\log\det \boldsymbol{X} - \operatorname{tr}(\boldsymbol{S}\boldsymbol{X}) - \lambda \|\boldsymbol{X}\|_1 \Big]$$

- Используя равенство $\|\boldsymbol{X}\|_1 = \max\limits_{\|\boldsymbol{U}\|_\infty \leqslant 1} \operatorname{tr}(\boldsymbol{X}\boldsymbol{U})$:

$$\max_{\boldsymbol{X} \text{ is PSD}} \min_{\|\boldsymbol{U}\|_\infty \leqslant \lambda} \Big[\log\det \boldsymbol{X} - \operatorname{tr}(\boldsymbol{X}, \boldsymbol{S} + \boldsymbol{U}) \Big]$$

- Можем поменять местами max и min и аналитически найти максимум внутренней функции $\log\det \boldsymbol{X} - \operatorname{tr}(\boldsymbol{X}, \boldsymbol{S} + \boldsymbol{U})$
- Результат: $\boldsymbol{X} = -\boldsymbol{S} - \boldsymbol{U}$, при этом $\operatorname{tr}(\boldsymbol{X}, \boldsymbol{S} + \boldsymbol{U}) = n$

Алгоритм [Banerjee et al., 2008] (трюки)

- Приходим к оптимизационной задаче:

$$\min_{\|\boldsymbol{U}\|_\infty \leqslant \lambda} \Big[-\log\det(\boldsymbol{S} + \boldsymbol{U}) - n \Big]$$

- Соответствующая задача с ограничениями ($\boldsymbol{W} = \boldsymbol{S} + \boldsymbol{U}$):

$$\widehat{\boldsymbol{\Sigma}} = \max\Big\{ \log\det \boldsymbol{W} : \|\boldsymbol{W} - \boldsymbol{S}\|_\infty \leqslant \lambda \Big\}$$

  называется дуальной задачей

- Задача гладкая и выпуклая по элементам $(\boldsymbol{W})_{ij}$
- Минус — высокая размерность задачи $n(n+1)/2$: для эффективного решения простыми оптимизационными методами должно быть выполнено $n \sim 1 \dots 10$ (например, ε-субоптимальный метод дает сложность $O(n^6 \log(1/\varepsilon))$)

Блочный координатный спуск

- Обозначим $\boldsymbol{A}_{\setminus k \setminus j}$ матрицу, полученную из $\boldsymbol{A}$ вычеркиванием j-той строки и k-того столбца
- Пусть k — номер обновления всей матрицы $\boldsymbol{W}$
- Алгоритм оптимизационного поиска $\boldsymbol{W}$: обновляем одну строку/столбец матрицы $\boldsymbol{W}$ на каждой итерации по строкам/столбцам $j = 1, \ldots, n$

Блочный координатный спуск

- Инициализация: установить $\boldsymbol{W}^{(0)} = \boldsymbol{S} + \lambda \boldsymbol{I}$

Блочный координатный спуск

- Инициализация: установить $\boldsymbol{W}^{(0)} = \boldsymbol{S} + \lambda \boldsymbol{I}$
- Итерации:
  1. Пусть $\boldsymbol{W}^{(j-1)}$ – текущая версия $\boldsymbol{W}$. Решим задачу (box-constrained quadratic program (QP))

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \boldsymbol{y}^\intercal \boldsymbol{W}^{(j-1)}_{\backslash j \backslash j} \boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{S}_j\|_\infty \leqslant \lambda \right\}$$

Блочный координатный спуск

- Инициализация: установить $\boldsymbol{W}^{(0)} = \boldsymbol{S} + \lambda \boldsymbol{I}$
- Итерации:
  1. Пусть $\boldsymbol{W}^{(j-1)}$ – текущая версия $\boldsymbol{W}$. Решим задачу (box-constrained quadratic program (QP))
  $$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \boldsymbol{y}^\intercal \boldsymbol{W}^{(j-1)}_{\setminus j \setminus j} \boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{S}_j\|_\infty \leqslant \lambda \right\}$$
  2. Обновим $\boldsymbol{W}^{(j)} \leftarrow \boldsymbol{W}^{(j-1)}$ с заменой строки $\boldsymbol{W}_j$ на $\widehat{\boldsymbol{y}}$

Блочный координатный спуск

- Инициализация: установить $\boldsymbol{W}^{(0)} = \boldsymbol{S} + \lambda \boldsymbol{I}$
- Итерации:
  1. Пусть $\boldsymbol{W}^{(j-1)}$ — текущая версия $\boldsymbol{W}$. Решим задачу (box-constrained quadratic program (QP))

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \boldsymbol{y}^\intercal \boldsymbol{W}^{(j-1)}_{\setminus j \setminus j} \boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{S}_j\|_\infty \leqslant \lambda \right\}$$

  2. Обновим $\boldsymbol{W}^{(j)} \leftarrow \boldsymbol{W}^{(j-1)}$ с заменой строки $\boldsymbol{W}_j$ на $\widehat{\boldsymbol{y}}$
- После пробега всех $j = 1, \ldots, n$ установим $\widehat{\boldsymbol{W}}^{(0)} \leftarrow \boldsymbol{W}^{(n)}$
- Критерий остановки: $\mathrm{tr}((\widehat{\boldsymbol{W}}^{(0)})^{-1} \boldsymbol{S}) - p + \lambda \|(\widehat{\boldsymbol{W}}^{(0)})^{-1}\|_1 \leqslant \varepsilon$

Результат алгоритма

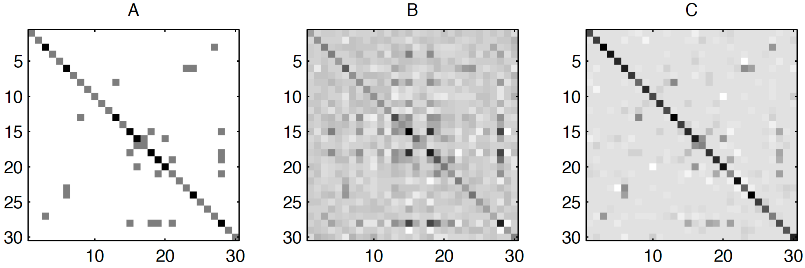


Рис.: A: разреженная матрица $\boldsymbol{\Sigma}^{-1}$;
B: оценка $\widehat{\boldsymbol{S}}^{-1}$ по выборке объема $\ell = 60$;
C: оценка $\widehat{\boldsymbol{\Sigma}}_{\ell_1}^{-1}$ с $\lambda = 0.1$.
Размерность задачи $n = 30$.

Связь λ и вероятности восстановления

- Рассмотрим истинную неизвестную графическую модель для заданного распределения
- Для заданного узла k пусть C_k — его компонента связности
- Пусть $\widehat{C}_k^\lambda$ — оценка его компоненты связности с параметром регуляризации λ
- **Утверждение.** Пусть параметр λ выбран так, что

$$\lambda(\alpha) = \max_{i>j}(\widehat{\sigma}_i\widehat{\sigma}_j)\frac{t_{n-2}(\alpha/2n^2)}{n-2+t_{n-2}(\alpha/2n^2)}.$$

Тогда вероятность того, что хотя бы одна компонента связности в графической модели будет восстановлена неверно не превосходит α:

$$\mathrm{P}(\exists k : \widehat{C}_k^\lambda \nsubseteq C_k) \leqslant \alpha).$$

Графическое Lasso

Графическое Lasso [Friedman et al., 2008]

- Можно заметить (см. дуальность норм), что задача на минимум

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \boldsymbol{y}^\mathsf{T} \boldsymbol{W}^{(j-1)}_{\setminus j \setminus j} \boldsymbol{y} : \|\boldsymbol{y} - \boldsymbol{S}_j\|_\infty \leqslant \lambda \right\}$$

имеет дуальную задачу

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left[\boldsymbol{y}^\mathsf{T} \boldsymbol{W}^{(j-1)}_{\setminus j \setminus j} \boldsymbol{y} - \boldsymbol{S}_j^\mathsf{T} \boldsymbol{y} + \lambda \|\boldsymbol{y}\|_1 \right]$$

- Обозначим $\boldsymbol{Q} = \left(\boldsymbol{W}^{(j-1)}_{\setminus j \setminus j} \right)^{1/2}$, $\boldsymbol{b} = \frac{1}{2} \boldsymbol{Q}^{-1} \boldsymbol{S}_j$, тогда

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left[\frac{1}{2} \|\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{b}\|_2^2 + \lambda \|\boldsymbol{y}\|_1 \right]$$

— известная как Lasso регуляризованная задача на минимум

Графическое Lasso: алгоритм

- Инициализация: установить $\boldsymbol{W} = \boldsymbol{S} + \lambda \boldsymbol{I}$
- Для каждого $j = 1, 2, \ldots, n, 1, 2, \ldots$ решить Lasso-задачу

$$\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left[\|\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{b}\|_2^2 + \lambda \|\boldsymbol{y}\|_1 \right],$$

  получив оценки $\widehat{\boldsymbol{y}}, \boldsymbol{w} = \boldsymbol{Q}^\intercal \boldsymbol{Q} \widehat{\boldsymbol{y}}$
- Заполнить j-ую строку матрицы $\boldsymbol{W}$ вектором $\boldsymbol{w}$
- Повторять до сходимости

Графическое Lasso: эксперименты

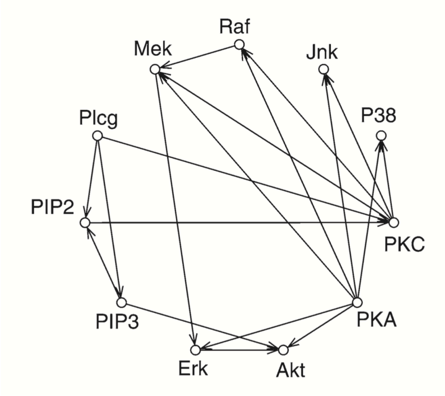


Рис.: Пример данных проточной цитометрии на $n = 11$ протеинах и $\ell = 7466$ клетках.
Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." Biostatistics 9.3 (2008): 432-441.

Графическое Lasso: эксперименты

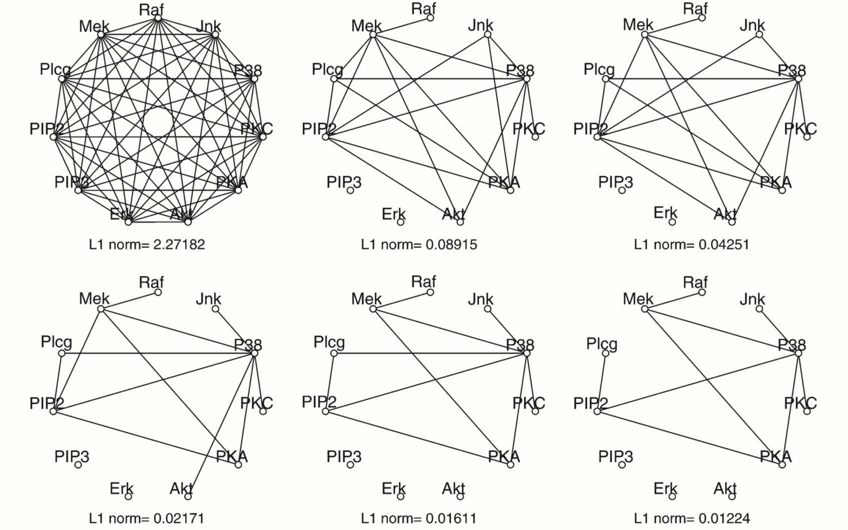


Рис.: Результаты применения графического Lasso к данным проточной цитометрии.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. ”Sparse inverse covariance estimation with the graphical

Насколько быстро граф. Lasso?

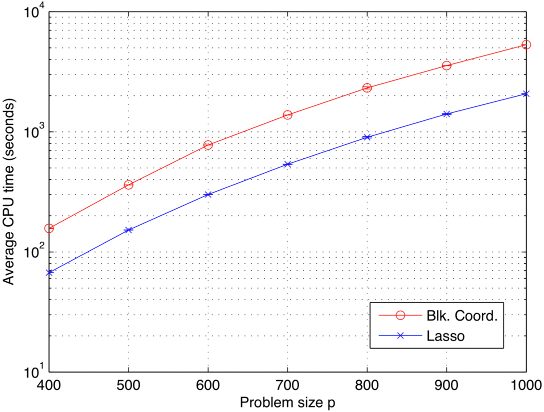


Рис.: Сравнение среднего времени работы двух оптимизационных алгоритмов: блочного координатного спуска и графического lasso.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. ”Sparse inverse covariance estimation with the graphical lasso.” Biostatistics 9.3 (2008): 432-441.

Современный подход: QUIC [Hsieh et al., 2014]

- QUIC: QUadratic approximation of Inverse Covariance matrices
- Метод второго порядка для решения ℓ_1-регуляризованных задач на ОМП в гауссовских предположениях
- Повторяются шаги Ньютоновского метода, использующие итеративную квадратичную оценку гауссовского NLL
- Специальный тип итераций (ускорение типа Armijo-rule) обеспечивает положительную определенность решения
- Идентификация блочной структуры для эффективных итераций
- Снижение сложности координатного спуска с $O(n^2)$ до $O(n)$
- Реализации на `R`, `MATLAB`, `scikit-learn`

Резюме лекции

- Иногда бывает необходимо оценивать коварицонные матрицы
- ℓ_1-регуляризация оценки помогает сделать оценки разреженными
- Блочно-координатный метод и графическое lasso — эффективные (хотя немножко устаревшие) алгоритмы оптимизации регуляризованного правдоподобия
- Более современные методы QUIC доступны в широко известных пакетах

Литература

- Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008).
  Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data.
  Journal of Machine learning research, 9(Mar):485–516.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008).
  Sparse inverse covariance estimation with the graphical lasso.
  Biostatistics, 9(3):432–441.
- Hsieh, C.-J., Sustik, M. A., Dhillon, I. S., and Ravikumar, P. (2014).
  Quic: quadratic approximation for sparse inverse covariance estimation.
  The Journal of Machine Learning Research, 15(1):2911–2947.