

Ответы на вопросы

К. Ф. Сафин

Московский физико-технический институт

18 января 2019 г.

Вопросы

- ▶ Методы понижения размерности. Метод главных компонент. Неотрицательные матричные разложения.
- ▶ Диаграмма Вороного линейных сегментов. Построение скелета многоугольной фигуры на основе диаграммы Вороного сегментов.
- ▶ Основная схема построения алгебраических расширений. Требования к семействам отображений.

Снижение размерности

Методы снижения размерности:

- ▶ Отбор признаков — перебор всех признаков
 - ▶ Add, AddDel
 - ▶ Lasso, линейная регрессия
- ▶ Синтез признаков — построение нового набора признаков
 - ▶ Метод главных компонент
 - ▶ t-SNE
 - ▶ Автоэнкодер

Метод главных компонент

Исходные признаки: $f_1(x), \dots, f_n(x)$

Новые признаки: $g_1(x), \dots, g_m(x)$, $m \leq n$

Линейное восстановление по новым признакам:

$$\hat{f}_j(x) = \sum_{s=1}^m g_s(x) u_{js}, \quad j = 1, \dots, n, \quad \forall x \in X$$

Точность восстановления:

$$\sum_{i=1}^l \sum_{j=1}^n \left(\hat{f}_j(x_i) - f_j(x_i) \right)^2 \rightarrow \min_{\{g_s(x_i)\}, \{u_{js}\}}$$

Метод главных компонент

$$F_{l \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_l) & \dots & f_n(x_l) \end{pmatrix}; \quad G_{l \times m} = \begin{pmatrix} g_1(x_1) & \dots & g_m(x_1) \\ \dots & \dots & \dots \\ g_1(x_l) & \dots & g_m(x_l) \end{pmatrix}$$

Матрица линейного преобразования:

$$U_{n \times m} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nm} \end{pmatrix}; \quad \hat{F} = GU^T \approx F$$

$$\sum_{i=1}^l \sum_{j=1}^n \left(\hat{f}_j(x_i) - f_j(x_i) \right)^2 = \|GU^T - F\|^2 \rightarrow \min_{G,U}$$

Теорема метода главных компонент

Если $m \leq \text{rk}F$, то минимум $\|GU^T - F\|^2$ достигается, когда столбцы U — с.в. матрицы $F^T F$, соответствующие m максимальным с.з. $\lambda_1, \dots, \lambda_m$, а матрица $G = FU$.

При этом:

- ▶ U — ортонормирована: $U^T U = I_m$,
- ▶ G — ортогональна: $G^T G = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$,
- ▶ $U\Lambda = F^T F U$; $G\Lambda = FF^T G$,
- ▶ $\|GU^T - F\|^2 = \|F\|^2 - \text{tr}\Lambda = \sum_{j=m+1}^n \lambda_j$.

Сингулярное разложение

Произвольная $l \times n$ -матрица представима в виде сингулярного разложения:

$$F = VDU^T$$

. Свойства:

- ▶ $V_{l \times n} = (v_1, \dots, v_n)$ ортогональна, $V^T V = I_n$, столбцы v_j — с.в. матрицы FF^T ,
- ▶ $U_{n \times n} = (u_1, \dots, u_n)$ ортогональна, $U^T U = I_n$, столбцы u_j — с.в. матрицы $F^T F$,
- ▶ $D_{n \times n}$ диагональна, $D = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$, $\lambda_j \geq 0$ — с.з. матриц $F^T F$ и FF^T .

Связь РСА и сингулярного разложения

При $m = n$:

- ▶ $\|GU^T - F\|^2 = 0$
- ▶ Представление $\hat{F} = GU^T = F$ точное и совпадает с сингулярным разложением при $G = V\sqrt{\Lambda}$:

$$F = GU^T = V\sqrt{\Lambda}U^T; \quad U^T U = I_m; \quad V^T V = I_m$$

- ▶ линейное преобразование U работает в обе стороны:

$$F = GU^T; \quad G = FU$$

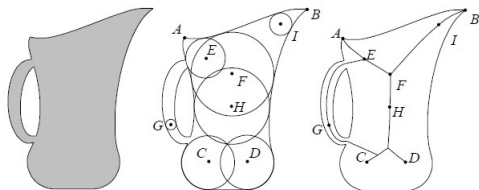
Т.к. новые признаки некоррелированы ($G^T G = \Lambda$), преобразование U — декоррелирующее.

Скелет фигуры

Пустой круг фигуры A — замкнутое множество точек

$$\tilde{S}_r(p) = \{q : q \in R^2, d(p, q) \leq r\} : \tilde{S}_r(p) \subset A$$

Скелет фигуры — множество центров всех её максимальных пустых кругов.



Структура скелета многоугольной фигуры

Типы ветвей скелета многоугольной фигуры:

- ▶ отрезки биссектрис,
- ▶ отрезки серединных перпендикуляров,
- ▶ ветви параболы.

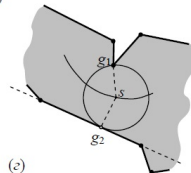
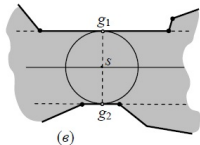
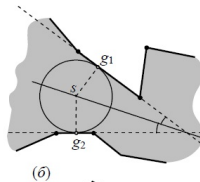
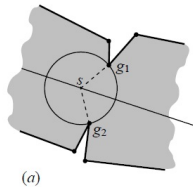
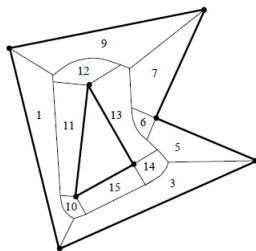
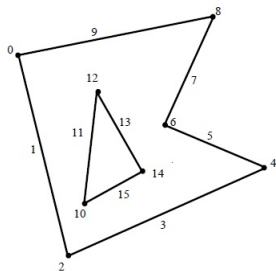


Диаграмма Вороного многоугольной фигуры

Сайт фигуры — вершина или сторона фигуры с инцидентными вершинами.

Ячейка Вороного сайта — г.м.т. фигуры, для которых этот сайт является ближайшим.

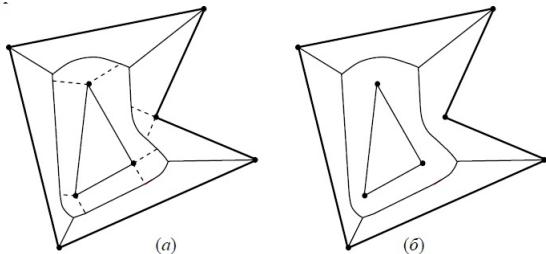
Диаграмма Вороного многоугольной фигуры — объединение границ смежных ячеек Вороного (бисекторов).



Получение скелета из диаграммы Вороного

Ветви скелета образуют все бисекторы пар сайтов, не являющихся соседними, и только они, т.к.:

- ▶ Скелет многоугольной фигуры является подмножеством её диаграммы Вороного.
- ▶ Бисектор пары сайтов, не являющихся соседними, содержится в скелете многоугольной фигуры.
- ▶ Бисектор пары соседних сайтов не содержится в скелете многоугольной фигуры.



Переборный алгоритм построения диаграммы Вороного

- ▶ Построение вершин диаграммы Вороного
 - ▶ Перебор всех троек сайтов.
 - ▶ Построение касательных окружностей для троек.
- ▶ Построение бисекторов диаграммы
 - ▶ Определение типа бисектора для смежных ячеек диаграммы.
 - ▶ Тип бисектора зависит от пары определяющих сайтов.

Задача синтеза корректного алгоритма

\mathfrak{I}_i — пространство начальных информации,

\mathfrak{I}_f — пространство конечных информации.

Отображение, реализуемое алгоритмом:

$$\mathfrak{I}_i \xrightarrow{A} \mathfrak{I}_f$$

Рассматриваются некорректные задачи: \exists решение, но не !

$$\mathfrak{m}^* = \{A | A : \mathfrak{I}_i \rightarrow \mathfrak{I}_f\}$$

I_{str} — структурная информация (требования к алгоритму),
описание подмножества допуст. отображений $\mathfrak{m}(I_{str}) \in \mathfrak{m}^*$.

\forall алгоритм, реализующий \forall допуст. отображение —
корректный.

Задача разрешима $\Leftrightarrow (I_{str}) \neq \emptyset$

Алгебраический подход

$\mathfrak{m}(\pi)$ — параметрическое семейство отображений.

Проблемы:

- ▶ $\mathfrak{m}(\pi) \cap \mathfrak{m}(I_{str}) = \emptyset$;
- ▶ $\mathfrak{m}(\pi) \cap \mathfrak{m}(I_{str}) \neq \emptyset$, но найти корректное отображение вычислительно сложно.

Алгебраический подход: корректирующие операции над \mathfrak{m}^* , т.е. вместо $A \in \mathfrak{m}^*$, строится $A \in f(\mathfrak{m}^*)$

Требование к f :

f — сводит оптимиз. задачу к известной, но не усложняет $\mathfrak{m}(\pi)$ настолько, чтобы поиск $extr$ стал невозможным.