

Занятия 6.2. Метод главных компонент

Гирдюк Дмитрий Викторович

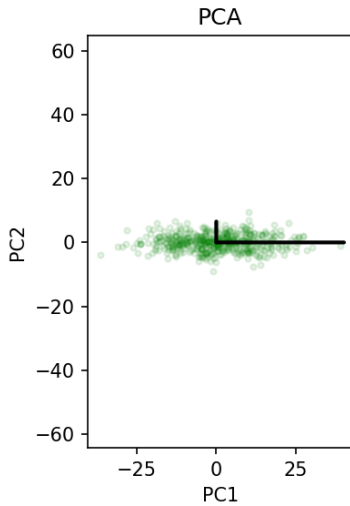
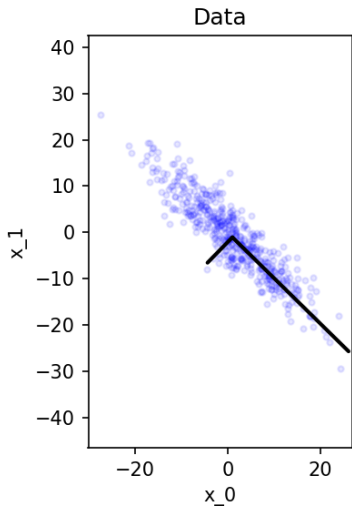
20 октября 2023 г.

СПбГУ, ПМ-ПУ, ДФС

Метод главных компонент

- Главными компонентами некоторого набора данных $X_{[m \times n]}$ является последовательность из n векторов, каждый из которых наилучшим образом (в смысле минимизации средних квадратов расстояний между наблюдениями и текущим вектором) подгоняется под данные, при этом каждый i -ый вектор ортогонален предыдущим $i - 1$ векторам.
- Новый ортогональный базис? Новый ортогональный базис!
- Метод главных компонент (Principal component analysis, PCA) – метод снижения размерности, основанный на построении набора из первых k таких ортогональных векторов.

РСА: визуализация



Общая теория 1

- Хотим получить такой нормированный ортогональный набор векторов $\mathbf{v}_j \in R^n, j = 1, 2, \dots, d$, которыми можно будет аппроксимировать исходные векторы $\mathbf{x}_i \in R^n, i = 1, 2, \dots, m$

$$\mathbf{x}_i \approx \sum_{j=1}^d c_{ij} \mathbf{v}_j, \quad i = 1, 2, \dots, m$$

- Нормируем (опционально шкалируем) данные!
- От простого к сложному: целевая функция (минимизация дисперсии/разброса проекции точек на главную компоненту) в случае $d = 1$

$$\frac{1}{m} \sum_{i=1}^m \|\mathbf{x}_i \perp \mathbf{v}\|_2^2 \longrightarrow \min_{\|\mathbf{v}\|_2=1}$$

Общая теория 2

- Теорема Пифагора позволяет преобразовать целевую функцию

$$\begin{aligned} \|x_i \perp v\|_2^2 + \langle x_i, v \rangle^2 &= \|x_i\|_2^2 \implies \\ \implies \frac{1}{m} \sum_{i=1}^m \langle x_i, v \rangle^2 &= \frac{1}{m} (Xv)^T (Xv) = \frac{1}{m} v^T X^T X v = \\ &= \frac{1}{m} v^T A v \longrightarrow \max_{\|v\|_2=1} \quad (1) \end{aligned}$$

- Симметричная матрица $A = X^T X$ есть ничто иное как ковариационная (учитывая шкалирование, еще и корреляционная) матрица. Важное свойство: собственные числа такой матрицы неотрицательны.
- Целевая функция для случай $d > 1$, проекция x_i на векторное подпространство $V = \{v_1, \dots, v_k\}$

$$\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^d \langle x_i, v_j \rangle^2 \longrightarrow \max_{V: \|v_j\|_2=1}$$

Общая теория 3

- Теперь рассмотрим разложение матрицы A

$$A = VDV^T$$

где матрица V есть ортогональная матрица, а D – диагональная.

- Заметим, что если матрица $A = \text{diag}(\lambda_1, \dots, \lambda_n)$ (пусть еще $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$), то максимум для (1) достигается для $v = e_1$

$$v^T A v = \sum_{j=1}^n \lambda_j v_j^2$$

- Отсюда следует, что для произвольной матрицы A , положив $v_1 = V e_1$, получим

$$v_1^T A v_1 = v_1^T V D V^T v_1 = e_1^T V^T V D V^T V e_1 = e_1^T D e_1 = \lambda_1$$

- Более того, для любого другого вектора \hat{v} , $\hat{v}^T A v \leq \lambda_1$.
- Для случая $d > 1$ все выводится аналогичным образом. Оптимальное решение – первые d столбцов матрицы V .
- Вопрос: как эту матрицу V искать?
- На самом деле матрица V в разложении $A = V D V^T$ есть ничто иное как матрица, столбцы которой являются собственными векторами матрица $A = X^T X$

$$A v_i = A V e_i = V D V^T V e_i = V D e_i = \lambda_i V e_i = \lambda_i v_i$$

- Последнее, что тут стоит отметить, это уникальность решения: если все собственные числа уникальны, то и разложение уникально. Если же встречаются кратные, то образуется целое подпространство собственных векторов, решающих задачу.

- Основанный на сингулярном разложении (Singular Value Decomposition, SVD).
- Итерационный алгоритм (Power Iteration).

Сингулярное разложение (SVD)

- Сингулярное разложение

$$X = USV^T$$

где $U_{[m \times m]}$ и $V_{[n \times n]}$ ортогональные матрицы размерности, а S – диагональная матрица размерности $m \times n$, значения на диагонали которой отсортированы в убывающем порядке.

- Столбцы матриц U и V называются левыми и правыми сингулярными векторами матрицы X соответственно.

Сингулярное разложение (SVD)

- Левые сингулярные векторы матрицы X есть ничто иное как собственные векторы матрицы XX^T . Аналогично, правые сингулярные векторы – собственные векторы матрицы $X^T X$.
- В самом деле

$$\begin{aligned} XX^T &= USV^T VS^T U^T = USS^T U^T = UD_1 U^T \Rightarrow \\ &\Rightarrow XX^T U = UD_1 \end{aligned}$$

$$\begin{aligned} X^T X &= VS^T U^T USV^T = VS^T SV = VD_2 V^T \Rightarrow \\ &\Rightarrow X^T X V = VD_2 \end{aligned}$$

- Отсюда следует алгоритм построения разложения: с помощью библиотек линейной алгебры построить спектральное разложения матриц XX^T и $X^T X$.

Итерационный алгоритм

Algorithm 1: Итерационный алгоритм поиска главной компоненты

input : Матрица $A = X^T X$

Выбираем произвольный нормированный вектор u_0 ;

for $i = 1, 2, \dots$ **do**

$u_i = A^i u_0$;

if $u_i / \|u_i\|_2 \approx u_{i-1} / \|u_{i-1}\|_2$ **then**

 Возвращаем u_i ;

-
- Важно отметить (доказывается по индукции):

$$A^{i+1} = A^i A = V D^i V^T V D V^T = V D^{i+1} V^T$$

- Как только нашли первую компоненту, проектируем данные ортогонально найденной главной компоненте, т.е. полагаем $x_i := x_i - \langle x_i, v_1 \rangle v_1$ и запускаем итерационный алгоритм заново.

Итерационный алгоритм: обсуждение

Почему последовательное домножение на вектор \mathbf{u}_0 матрицы \mathbf{A} приводит нас в конечном итоге к главной компоненте?

- $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^T$
- $\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{D}$
- Учитывая то, что по столбцам у матрицы \mathbf{V} стоят собственные векторы, формирующие ортогональный базис, то любой вектор в этом базисе может быть расписан как $\mathbf{u}_0 = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$.
- Тогда получаем

$$\begin{aligned}\mathbf{A}^i\mathbf{u}_0 &= c_1\mathbf{A}^i\mathbf{v}_1 + \dots + c_n\mathbf{A}^i\mathbf{v}_n = c_1\lambda_1^k\mathbf{v}_1 + \dots + c_n\lambda_n^k\mathbf{v}_n = \\ &= \lambda_1^k(c_1\mathbf{v}_1 + c_2\frac{\lambda_2^k}{\lambda_1^k}\mathbf{v}_2 + \dots + c_n\frac{\lambda_n^k}{\lambda_1^k}\mathbf{v}_n)\end{aligned}$$

- Откуда следует, что при $i \rightarrow \infty$, учитывая $\lambda_1 \geq \dots \geq \lambda_n$,
 $\mathbf{A}^i\mathbf{u}_0 \rightarrow C(i)\mathbf{v}_1$

- Простая интерпретация метода главных компонент: компоненты последовательно выбираются так, чтобы дисперсия проекции данных на нее была максимальной. Следует это из центрированности данных и вида целевой функции (1).
- Выбираем число главных компонент на основе объясненной дисперсии, равной кумулятивной сумме отнормированных собственных чисел, соответствующих собственным векторам (главным компонентам). Или используем правило Кайзера:

$$\lambda_i > \frac{1}{n} \text{trace} \mathbf{A}$$

- Еще раз, нормализуем (и шкалируем) данные!
- Интерпретация главных компонент зачастую затруднительна.

- Нелинейная структура в данных – используйте другой метод (рассмотрим позже в курсе).
- Например, Kernel PCA! Все отличие лишь в том, что находим собственные векторы не ковариационной матрицы $X^T X$, а ядровой матрицы $K : k_{ij} = \kappa(x_i, x_j)$.
- Самое главное: $Y = XV$. Но мы можем взять лишь первые d компонент для аппроксимации: $Y_k = XV_k$.