

Доказательство связи РСА и собственных векторов ковариационной матрицы

Черников Кирилл, Вдовина Анастасия

Содержание

1	Центрирование данных и ковариация	2
2	Свойства ковариационной матрицы и её разложение	2
3	Разложение направления и выражение дисперсии	3
4	Оптимальный выбор первого направления	3
5	Выбор первых k компонент	3
6	Доля объяснённой дисперсии	4
7	Пример: матрица 2×2	4

1 Центрирование данных и ковариация

Предварительно устраняем влияние смещения за счёт разных средних значений признаков. Для этого центрируем матрицу данных $X \in \mathbb{R}^{n \times m}$ по столбцам:

$$X_c = X - \bar{X}, \quad (X_c)_{ij} = X_{ij} - \frac{1}{n} \sum_{k=1}^n X_{kj}.$$

Теперь у каждого столбца среднее равно нулю, и дальнейшие расчёты дисперсии будут корректными.

По центрированным данным определим матрицу ковариаций:

$$\Sigma = \frac{1}{n-1} X_c^\top X_c.$$

Эта матрица отражает попарное поведение признаков: по диагонали стоят дисперсии, вне диагонали — ковариации. Так мы получили объект, описывающий, как признаки «двигаются вместе».

2 Свойства ковариационной матрицы и её разложение

Матрица Σ обладает двумя ключевыми свойствами:

- **Симметричность:** $\Sigma^\top = \Sigma$. Это следует из выражения $X_c^\top X_c$.
- **Положительная определённость.**

По спектральной теореме любая вещественная симметричная матрица допускает диагонализацию ортонормированным преобразованием. Значит, существуют собственные векторы v_1, \dots, v_m и неотрицательные собственные значения $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, такие что

$$\Sigma = V \Lambda V^\top, \quad V = [v_1, \dots, v_m], \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m).$$

Это разложение раскрывает «главные оси» действия матрицы ковариаций.

3 Разложение направления и выражение дисперсии

Любой единичный вектор w ($\|w\| = 1$) можно разложить по собственным векторам:

$$w = \sum_{i=1}^m c_i v_i, \quad \sum_{i=1}^m c_i^2 = 1, \quad c_i = v_i^\top w.$$

Проекция всех точек на w задаётся $X_c w$, а её дисперсия вычисляется как

$$w^\top \Sigma w.$$

Подставляя разложение, получаем

$$w^\top \Sigma w = \sum_{i=1}^m c_i^2 \lambda_i.$$

Такой вид подчёркивает, что каждая компонента v_i вносит в общую дисперсию вклад $c_i^2 \lambda_i$.

4 Оптимальный выбор первого направления

Чтобы найти направление максимальной дисперсии, нужно максимизировать сумму $\sum_i c_i^2 \lambda_i$ при условии $\sum_i c_i^2 = 1$. Поскольку λ_1 — наибольшее, все весовые коэффициенты c_i^2 должны «сосредоточиться» на $i = 1$. Таким образом

$$w_1 = v_1, \quad \max_{\|w\|=1} w^\top \Sigma w = \lambda_1.$$

Именно первый собственный вектор даёт наибольшую дисперсию.

5 Выбор первых k компонент

Если требуется сохранить k измерений, выбираем ортонормированный набор $W = [w_1, \dots, w_k]$ так, что $W^\top W = I_k$ и максимизируем

$$\text{Tr}(W^\top \Sigma W) = \sum_{i=1}^k w_i^\top \Sigma w_i.$$

Аналогично предыдущему рассуждению оптимальным оказывается выбор первых k собственных векторов:

$$W = [v_1, \dots, v_k], \quad \max \text{Tr}(W^\top \Sigma W) = \sum_{i=1}^k \lambda_i.$$

Так обеспечивается максимальное суммарное сохранение дисперсии.

6 Доля объяснённой дисперсии

Оценим, какую часть от общей дисперсии захватывают первые k компонент:

$$\gamma(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i}.$$

Обычно выбирают k так, чтобы $\gamma(k) \geq 0.9$ или 0.95 , гарантируя сохранение 90–95% вариации.

7 Пример: матрица 2×2

Для наглядности рассмотрим

$$\Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}.$$

Характеристическое уравнение даёт $\lambda_{1,2} = (7 \pm \sqrt{17})/2 \approx 5.56, 1.44$, собственные векторы $v_1 \approx (0.79, 0.61)^\top$, $v_2 \approx (-0.61, 0.79)^\top$. Таким образом первое направление максимально рассеивает точки (λ_1), второе — следующий вклад (λ_2).

Заключение

В результате мы показали, что метод PCA напрямую вытекает из спектрального анализа ковариационной матрицы Σ . В частности:

- Направления максимальной дисперсии — собственные векторы v_i .
- Их важность задаётся соответствующими λ_i .
- Отбор первых k векторов сохраняет максимальную суммарную дисперсию $\sum_{i=1}^k \lambda_i$.
- PCA — это формализация идеи «главных осей» через разложение $\Sigma = V\Lambda V^\top$.

Также показали, что оптимальная проекция данных на k измерений реализуется через первые k собственных векторов ковариационной матрицы Σ , обеспечивая максимальное сохранение дисперсии при снижении размерности.