Доказательство совпадения направлений PCA с собственными векторами

Доказательство. **Дано:**

- Матрица данных $X \in \mathbb{R}^{d \times n}$ (d признаков, n наблюдений)
- Данные центрированы: $\sum_{i=1}^{n} x_i = 0$
- Ковариационная матрица: $C = \frac{1}{n} X X^T$

Требуется доказать: Направления главных компонент являются собственными векторами матрицы C.

Доказательство для первой компоненты:
1. Задача оптимизации:

$$\max_{\|w\|=1} w^T C w \tag{1}$$

2. Функция Лагранжа с ограничением $\|w\|^2 = 1$:

$$\mathcal{L}(w, \lambda) = w^T C w - \lambda(w^T w - 1) \tag{2}$$

3. Условия оптимальности:

$$\frac{\partial \mathcal{L}}{\partial w} = 2Cw - 2\lambda w = 0 \tag{3}$$

4. Упрощаем:

$$Cw = \lambda w \tag{4}$$

5. Полученное уравнение:

- Является задачей на собственные значения
- w — собственный вектор матрицы C
- λ — соответствующее собственное значение

Обобщение для k-й компоненты:
Для последующих компонент задача имеет вид:

$$\max_{\substack{\|w\|=1 \\ w \perp w_1, \ldots, w \perp w_{k-1}}} w^T C w \tag{5}$$

Решение аналогично приводит к:

$$Cw_k = \lambda_k w_k \tag{6}$$

Вывод:

- Оптимальные направления PCA совпадают с собственными векторами C
- Собственные векторы $\{w_1, \ldots, w_d\}$ образуют ортонормированный базис
- Собственные значения $\lambda_1 \geq \ldots \geq \lambda_d$ соответствуют дисперсиям

□