# Метод главных компонент
# Principal Component Analisys, PCA

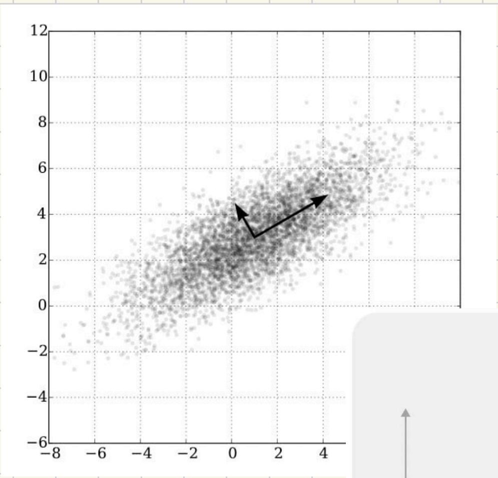
## Суть метода.

Даны центрированные данные $X \in \mathbb{R}^{n \times d}$ ← кол-во признаков

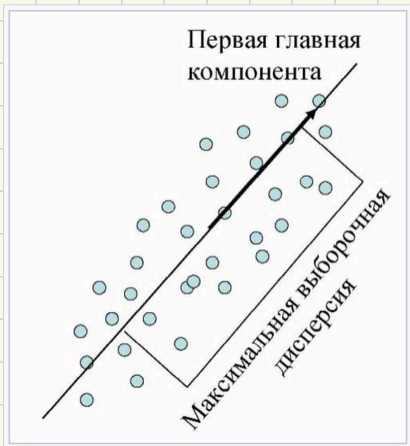
Мы хотим заменить множество коррелированных признаков меньшим числом новых переменных — главных компонент — таких, чтобы:

- первая компонента улавливала максимально возможную дисперсию данных
- вторая — максимальную из оставшейся при ортогональности к первой, и так далее для последующих компонент.

Геометрически: ищем k-размерное подпространство с ортонорм. базисом, на которое ортопроекция $X$ „теряет" как можно меньше информации








Первая главная компонента максимизирует выборочную дисперсию проекции данных

# Формальная постановка задачи. (и решение)

Пусть дана матрица данных $X \in \mathbb{R}^{n \times d}$ — $n$ наблюдений по $d$ признакам. Предполагается центровка по столбцам: $\frac{1}{n} \sum_{i=1}^{n} x_{ij} = 0 \quad \forall j \in \{1, 2, \dots, d\}$

Хотим перейти к $Z \in \mathbb{R}^{n \times l}$, где $l < d$

Ищем главные компоненты $u_1, u_2, \dots, u_d \in \mathbb{R}^d$.

Требования: 1) $\|u_i\|^2 = 1$ 2) $\langle u_i, u_j \rangle = 0$, $i \neq j$

3) при проецировании $X$ на $u_1, u_2, \dots, u_l$ получается макс. дисперсия проекций среди всех способов выбрать

- Если $x$ проецируем на $u_i$: $u_i(\underbrace{u_i^T u_i}_{1})^{-1} u_i^T x = u_i \langle u_i, x \rangle$

Был $x$ $\xrightarrow[l \text{ компонент}]{\text{проецируем на 1-е}}$ $\begin{array}{c|c} \langle u_1, x \rangle & u_1 \\ \langle u_2, x \rangle & u_2 \\ \vdots & \\ \langle u_l, x \rangle & u_l \end{array}$ $\Rightarrow$ Воспринимаем $x$ новый как набор коэф-в $\underbrace{\langle u_1, x \rangle, \dots, \langle u_l, x \rangle}_{\text{новые признаки}}$

ортонорм. базис

новые признаки (лин. комб. исходных)

$\Uparrow$

! $\langle u_1, x \rangle = \sum_{j=1}^{d} u_{1j} \cdot x_j$ — т.е взвеш. сумма исходных признаков

Найдём сначала $u_1$ (помним про max. дисперсию)

$$\begin{cases} \|X u_1\|^2 \longrightarrow \max_u \\ \|u_1\|^2 = 1 \end{cases}$$

$X u_1$ — вектор скалярнах произведений

Вспомним (узнаем?) метод множителей Лагранжа

$L(x, \lambda) = f(x) + \lambda \cdot \varphi(x)$

Вернёмся к нашей задаче

$L(u_1, \lambda) = \|X u_1\|^2 + \lambda (\|u_1\|^2 - 1)$

! $\|X u_1\|^2 = u_1^T X^T X u_1 \qquad \|u_1\|^2 = u_1^T u_1$

Теперь надо продифференцировать по $u_1$

Быстрый экскурс:

$$f(x_0+h) = f(x_0) + [D_{x_0} f](h) + o(\|h\|)$$

$$f(u_1) = \|Xu_1\|^2 = u_1^T X^T X u_1 = \langle Xu_1, Xu_1 \rangle$$

$$[D_{u_1} \langle Xu_1, Xu_1 \rangle](h) = \langle [D_{u_1}(Xu_1)](h), Xu_1 \rangle + \langle Xu_1, [D_{u_1}(Xu_1)](h) \rangle$$

$$= 2\langle Xu_1, [D_{u_1}(Xu_1)](h) \rangle = 2\langle Xu_1, Xh \rangle =$$

$$= 2\langle X^T X u_1, h \rangle \Rightarrow \nabla_{u_1} f = 2X^T X u_1$$

Аналогично с $f(u_1) = \|u_1\|$ $\quad \nabla_{u_1} f = 2u_1$

$$\Rightarrow \nabla_{u_1} L = 2X^T X u_1 + 2\lambda u_1 = 0 \qquad X^T X u_1 = -\lambda u_1 \quad -\lambda = \mu$$

$\Rightarrow u_1$ – собственный вектор $X^T X$

Вернёмся к формуле $\|Xu_1\|^2 = u_1^T X^T X u_1 = \mu u_1^T u_1 = \mu$

(т.е. мы максимизируем собств. значение)

$u_2$ – второй собств. вектор по 2-ому максимальному собственному значению

$\vdots$ $\qquad \Rightarrow$ тогда $\underset{n\times d}{Z} = X \cdot \underset{d\times \ell}{U}$

$u_\ell$ так далее

---

$$f(x_0+h) = \|Xu_1 + Xh\|^2 = \langle Xu_1 + Xh, Xu_1 + Xh \rangle$$

$$\overset{\|Xu_1\|^2 = f(x_0)}{u_1^T X^T X u_1} + \underbrace{u_1^T X^T X h + h^T X^T X u_1}_{2\langle X^T X u_1, h\rangle} + \overset{\|Xh\|^2 = o(h)}{h^T X^T X h}$$

Считаем по определению (в лоб)

## Связь с SVD

Мат. содержание метода глав. компонент - спектральное разложение ковариац. матрицы $C$ ($C = X^T X$, но мы не знаем (вроде?) что это, поэтому устно немного инфы), то есть представление прост-ва данных в виде суммы взаимно орт. собств. подпростр. $C$

$$C = X^T X = Q \cdot \Lambda \cdot Q^{-1}$$

$\Lambda$ - диагональная матрица собств. значений

$Q$ - матрицы, столбцы собств. вектора

! $X^T X$ - симм. + веществ. $\Rightarrow$ $C = Q \cdot \Lambda \cdot Q^T$

$$X = U \Sigma V^T$$

$$X^T X = (U \Sigma V^T)^T (U \Sigma V^T) = V \Sigma^T U^T U \Sigma V^T = V \Sigma^T \Sigma V^T$$

$$\Lambda = \Sigma^T \Sigma \qquad (\lambda_i = \sigma_i^2)$$

## Как выбирать число компонент?

- Кумулятивная доля объяснённой дисперсии - выбираем min. $k$, где $\sum_{i=1}^{k} \lambda_i : \sum_{i=1}^{d} \lambda_i \geq \tau$
(τ - процент, который хотим сохранить, например 0,9)

- Метод „сломанной трости" (scree plot) - строим график собственных значений и их доли. Ищем „излом" - точку, где значения резко уменьшаются.
Математически.
$C$ - выборочная ковариационная матрица
$\lambda_i$ - собственное значение
$l_i$ - ожидаемая доля дисперсии для j-й компоненты „случайных" данных без структуры
$l_i = \frac{1}{d} \cdot \sum_{j=i}^{d} \frac{1}{j}$ (пример $l_3 = \frac{1}{6} \cdot \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right)$ )
Сохраняем вектор, если $\frac{\lambda_i}{\mathrm{tr}\, C} > l_i$ (и все до $i$ тоже)

- Правило Кайзера - значимы главные компоненты, для которых $\lambda_i > \frac{1}{d} \cdot \mathrm{tr}\, C$
(хорошо работает на простых случаях)

- По числу обусловленности. Выбираем $\kappa_0 > 1$ и оставляем компоненты, для которых $\lambda_i > \frac{\lambda_1}{\kappa_0}$

## Примеры применения.

- Сжатие изображений
- Визуализация (при снижение размерности до 2-3)
- Финансы (факторы доходностей)

- Подавление шума на изображениях
- Биоинформатика