Метод главных компонент

- один из основных методов уменьшения размерности данных с потерей наименьшего количества информации

Постановка задачи:

Пусть $x_i = (f_1(x_i), .., f_n(x_i))$, $i = \overline{1, l}$ — признаковое описание объектов обучающей выборки, $f_j(x)$, $j = \overline{1, n}$ — числовые признаки

Рассмотрим матрицу F:

$$F_{l \times n} = \begin{bmatrix} f_1(x_1) & .. & f_n(x_1) \\ & \cdots & \\ f_1(x_l) & .. & f_n(x_l) \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_l \end{bmatrix}$$

Пусть $z_i = (g_1(x_i), .., g_m(x_i))$ — признаковое описание тех же объектов, $z \in R^m$, $m < n$

$$G_{l \times m} = \begin{bmatrix} g_1(x_1) & .. & g_m(x_1) \\ & \cdots & \\ g_1(x_l) & .. & g_m(x_l) \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_l \end{bmatrix}$$

Пусть U: $\hat{x} = zU^T$

$$\Delta^2(G, U) = \sum_{i=1}^{l} \|\hat{x}_i - x_i\|^2 = \sum_{i=1}^{l} \|z_i U^T - x_i\|^2 = \|GU^T - F\|^2 \to \min_{G, U}$$

$$\operatorname{rg} G = \operatorname{rg} U = m \leq \operatorname{rg} F$$

Т Если $m \leq \operatorname{rg} F$, то минимум $\Delta^2(G, U)$ достигается, когда столбцы матрицы U есть с.в. $F^T F$, соответствующие m максимальным с.з.
При этом $G = FU$, U и G — ортогональны

$\blacktriangleright$ $\underline{f - fo}$:

$$\begin{cases} \frac{\partial \Delta^2}{\partial G} = (GU^T - F)U = 0 \\ \frac{\partial \Delta^2}{\partial U} = G^T(GU^T - F) = 0 \end{cases}$$ т.к G и U — матрицы полного ранга $\Rightarrow$ $\begin{cases} G = FU(U^TU)^{-1} \\ U = F^TG(G^TG)^{-1} \end{cases}$

R: $GU^T = (GR)(R^{-1}U^T)$, R подбираем так, что U^TU и G^TG — диагональные

$G^TG = \Lambda = \operatorname{diag}(\lambda_1, .., \lambda_m)$

$U^TU = I_m$

$\Rightarrow \begin{cases} G = FU \\ U\Lambda = F^TG \end{cases}$ $\Rightarrow$ $U\Lambda = F^TFU$, т.е столбцы U — с.в матрицы F^TF, диагональ $\lambda_1, .., \lambda_m$ — соответствующие им с.з. (при таком соотношении диагонали)

$$\Rightarrow \Delta^2(G, U) = \|F - GU^T\|^2 = \operatorname{tr}(F^T - UG^T)(F - GU^T) = \operatorname{tr} F^T(F - GU^T) = \operatorname{tr} F^TF - \operatorname{tr} F^TGU^T =$$
$$= \|F\|^2 - \operatorname{tr} U\Lambda U^T = \|F\|^2 - \operatorname{tr}\Lambda = \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{m} \lambda_j - \sum_{j=m+1}^{n} \lambda_j$$

где $\lambda_1, .., \lambda_n$ — все с.з. матрицы F^TF

минимум Δ^2 достигается, когда $\lambda_1, .., \lambda_m$ — наибольшие из с.з. и с.в $u_1, .., u_m$, отвечающие им с.з. — главные компоненты

Геометрический смысл:

Метод аппроксимирует n-мерное облако наблюдений до m-мерного эллипсоида, полуоси которого будут иметь главными компонентами.

SVD-разложение

- декомпозиция комплексной матрицы с целью приведения ее к каноническому виду.

Т $\exists A$ - произвольная матрица из $M_{m,n}$

$\exists$ ОНБ $\{e^k\}_{k=1}^n \subset \mathbb{C}^n$, $\{q^k\}_{k=1}^m \subset \mathbb{C}^m$ и $\tilde\sigma_1 \geq \tilde\sigma_2 \geq .. \geq \tilde\sigma_r$, $0 \leq r \leq \min\{m,n\}$, $\tilde\sigma_i > 0$, $i = \overline{1,r}$:

$$Ae^k = \begin{cases} \tilde\sigma_k q^k, & k \leq r \\ 0, & k > r \end{cases} \quad (1)$$

числа $\tilde\sigma_1, \tilde\sigma_2, .., \tilde\sigma_r$ - сингулярные числа матрицы A.

базисы $\{e^k\}_{k=1}^n$, $\{q^k\}_{k=1}^m$ - сингулярные базисы матрицы A

r - размерность $\mathrm{Im}(A)$, т.е. $r = \mathrm{rg}(A)$

Док-во:

Матрица A^*A самосопр. и неотрицательна

$((A^*A)^* = A^*A, (A^*Ax, x) = (Ax, Ax) \geq 0 \quad \forall x \in \mathbb{C}^n)$

$\Rightarrow \exists$ ОНБ с.в. $\{e^k\}_{k=1}^n$ матрицы A^*A. Все с.ч. ≥ 0

$$A^*Ae^k = \tilde\sigma_k^2 e^k, \quad k = \overline{1,n}, \quad \tilde\sigma_k^2 \geq 0, \quad k = \overline{1,n}$$

$\exists\ \tilde\sigma_1 \geq \tilde\sigma_2 \geq .. \geq \tilde\sigma_r > 0$, $\tilde\sigma_{r+1} = .. = \tilde\sigma_n = 0$

$\exists\ z^k = Ae^k$, $k = \overline{1,r}$

$$(z^p, z^q) = (Ae^p, Ae^q) = (A^*Ae^p, e^q) = \tilde\sigma_p^2 (e^p, e^q)$$

$$\Rightarrow (z^p, z^q) = \begin{cases} 0, & p \neq q \\ \tilde\sigma_p^2, & p = q \end{cases}$$

$\Rightarrow$ векторы $q^k = \tilde\sigma_k^{-1} Ae^k$, $k = \overline{1,r}$ образуют ОН систему в $\mathbb{C}^m$

(если $r < m$, дополним ее произвольными векторами q^k, $k = \overline{r+1, m}$ до ОНБ $\mathbb{C}^m$)

Матричное представление:

равенство (1) часто записывают в виде $A = V\Sigma W^*$

V - матрица со столбцами $\{q^k\}_{k=1}^m$

W - -"- $\{e^k\}_{k=1}^n$

Σ - диаг. матрица:

$$\Sigma = \begin{bmatrix} \tilde\sigma_1 & & & \\ & \ddots & & 0 \\ & & \tilde\sigma_r & \\ 0 & & & 0 \end{bmatrix}$$

Геометрический смысл:

$\exists A$ - линейный оператор. Его можно разбить на более простые операторы вращения, растяжения. Компоненты сингулярного разложения показывают эти преобразования