

SVD i metoda najmniejszych kwadratów

AB

19 kwietnia 2017

SVD - *singular-value decomposition*;

Metoda rozwiązywania problemów (wartości własne rzeczywistej macierzy symetrycznej, rozwiązywanie układu równań liniowych, najmniejsza odległość w metodzie najmniejszych kwadratów, ...) jednocześnie diagnozująca problem.

W szczególności, bardzo różne co do wielkości tzw. wartości singularne wskazują na liniowe zależności między danymi. Liniowość prowadzi do problemów obliczeniowych ponieważ małe zmiany w danych prowadzą do dużych zmian w wynikach.

Rozpatrzmy 3 płaskie wektory:

$$A = (1, 0)^T$$

$$B = (1, 0.1)^T$$

$$C = (0.95, 0.1)^T$$

Wektory C i B są bardzo podobne. Tworzą z wektorem A w przybliżeniu kąt 6° . Jednak kiedy kąty między $A + B$ i $A + C$ są prawie takie same, bliskie 0.07° to kąt między $A - B$ i $A - C$ wynosi 26.57° .

Z drugiej strony, wektory

$$A = (1, 0)^T$$

$$D = (0, 1)^T$$

$$E = (0, 0.95)^T$$

dają dla $A + D$ i $A + E$ oraz $D - A$ i $E - A$ kąty równe w przybliżeniu 1.5° .

Podsumowanie

Suma wektorów *kolinearnych* jest dobrze określona, ich różnica nie jest.

Suma i różnica wektorów *niekolinearnych* jest dobrze określona.

Praca z macierzami — dekompozycje macierzy.

Dekompozycja QR

$A = QR$, $A(m \times n)$, $Q^T Q = Q Q^T = 1_m$, $R_{ij} = 0$ dla $i > j$.

Dekompozycja SVD

Jeśli $R = SV$, gdzie $S(m \times n)$, $S_{ij} = 0$ dla $i \neq j$, $V(n \times n)$, $V^T V = V V^T = 1_n \longrightarrow A = QRS$.

Ponieważ zera poniżej diagonalii R i S powodują, że kolumny $m + 1, \dots, n$ macierzy Q są dowolne, więc można je opuścić, otrzymując macierz $U(m \times n)$ taką, że $U^T U = 1_n$. Opuszczenie wierszy $n + 1, \dots, m$ obu macierzy R i S pozwala zapisać rozkład w postaci $A = UR = USV^T$, gdzie $A(m \times n)$, $U(m \times n)$ a $U^T U = 1_n$, $R(n \times n)$ jest macierzą górną trójkątną, $S(n \times n)$ jest diagonalna, a $V(n \times n)$ jest ortogonalna.

Dekompozycja LU

$A(n \times n) = LU$ gdzie L - dolna macierz trójkątna, U - górna macierz trójkątna.

Dekompozycja Choleskiego

Jeśli A jest symetryczna to $A = LL^T$. (Postać skalowana: $A = LDU$, gdzie D - diagonalna.)

Zadanie

Pokazać, że jeśli $A = USV^T$ to macierz $A^+ = VS^+U^T$, gdzie $S_{ii}^+ = 1/S_{ii}$ dla $S_{ii} \neq 0$ i $S_{ii}^+ = 0$ dla $S_{ii} = 0$ spełnia równości:

$$AA^+A = A, \quad (AA^+)^T = AA^+, \quad A^+AA^+ = A^+, \quad (A^+A)^T = A^+A.$$

Problem: Znaleźć macierz ortogonalną $V(n \times n)$, która transformuje macierz $A(m \times n)$ w macierz B o ortogonalnych kolumnach b_i , tzn.

$$B = AV = (b_1, b_2, \dots, b_n) \quad (1)$$

gdzie

$$b_i^T b_j = S_i^2 \delta_{ij} \quad (2)$$

oraz

$$VV^T = V^T V = \mathbf{1}_n. \quad (3)$$

Przyjmujemy, że S_i są dodatnie i będziemy je nazywać wartościami singularnymi macierzy A .

Wektory

$$u_j = b_j S_j \quad (\text{niezerowe}) \quad (4)$$

są jednostkowymi wektorami ortogonalnymi. Zbierając u_j w macierz $m \times n$, a S_j w macierz diagonalną S można zapisać

$$B = US \quad (5)$$

gdzie

$$U^T U = 1_n. \quad (6)$$

Dla pozostałych, zerowych wartości S_j , konstruujemy wektory u_j ortogonalne do pozostałych lub przyjmujemy je jako bardzo małe. Kombinacje (1) i (6) daje

$$AV = US, \quad \text{lub} \quad A = USV^T, \quad (7)$$

co jest rozkładem SVD macierzy A .

- Obroty
- Metoda Givensa

Problem. Rozwiązać układ równań $Ax = b$ metodą najmniejszych kwadratów.

Rozwiązanie. Zdefiniować wektor reszt $r = b - Ax$ i znaleźć takie x by wektor r był minimalny w sensie normy $\|r\|_E = (r^T r)^{1/2} = \sqrt{\sum_i r_i^2}$. SVD dla A pozwala zapisać rozwiązanie równania $Ax = b$ w postaci

$$x = VS^+U^Tb, \quad (8)$$

gdzie $S_i^+ = 1/S_i$ dla $S_i \neq 0$ i $S_i = 0$ w przypadku $S_i = 0$. Tak zdefiniowana S^+ jest niepraktyczna numerycznie (zero jest rzadkim wynikiem...) Przyjmijmy, że

$$S_i^+ = \begin{cases} 1/S_i, & \text{dla } S_i > q, \\ 0, & \text{dla } S_i \leq q, \end{cases} \quad (9)$$

gdzie q jest zadane przez użytkownika.

Permitting S^+ to depend on a user-defined tolerance places upon him/her the responsibility for deciding the degree of linear dependence in his/her data. In an economic modelling situation, for instance, columns of U corresponding to small singular values are almost certain to be largely determined by errors or noise in the original data. On the other hand, the same columns when derived from the tracking of a satellite may contain very significant information about orbit perturbations. Therefore, it is not only difficult to provide an automatic definition for S^+ , it is inappropriate.

J C Nash

Znaleźć SVD macierzy

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Obliczamy $A^T A$

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

i wartości własne $A^T A$

$$|A^T A - \lambda I| = \det \begin{vmatrix} 4 - \lambda & 2 \\ 2 & 4 - \lambda \end{vmatrix} = (2 - \lambda)(6 - \lambda)$$

Stąd $\sqrt{6}$ i $\sqrt{2}$ są wartościami singularnymi macierzy A . Wektory własne związane z wartościami własnymi 6 i 2 są równe

$$v_1 = (1/\sqrt{2}, 1/\sqrt{2})^T, \quad v_2 = (1/\sqrt{2}, 1/\sqrt{2})^T.$$

Dalej

$$AA^T = \begin{bmatrix} 2 & 2 & 2 & 0 \\ 2 & 2 & 2 & 0 \\ 2 & 2 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Wektory własne jednostkowe należące do wartości własnych $\lambda = 6$ i $\lambda = 2$ są

$$u_1 = 1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}, 0)^T, \quad u_2 = (0, 0, 0, 1)^T.$$

Aby uzupełnić u_1 i u_2 do bazy ortonormalnej w R^4 ...

Płaszczyzna $x + y + z = 0$ jest zbiorem wektorów w R^3 ortogonalnych do wektora $n = (1, 1, 1)^T$. Jednym z takich wektorów jest $\sigma_1 = (1, -1, 0)^T$. Inny można wyznaczyć z iloczynu wektorowego $\sigma_2 = \sigma_1 \times n = (-1, -1, 2)$. Normalizując oba wektory i zanurzając je w R^4 (zerowa 4-ta składowa) otrzymamy bazę ortonormalną R^4 .

SVD macierzy A jest $A = USV^T$, gdzie

$$U = \begin{bmatrix} 1/\sqrt{3} & 0 & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 0 & 2/\sqrt{6} \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$S = \begin{bmatrix} \sqrt{6} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 1/\sqrt{2} & \sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Sprawdzić.

Znaleźć SVD macierzy $A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$

Odpowiedź: $U = [(1/\sqrt{2}, -1/\sqrt{2}), (1/\sqrt{2}, 1/\sqrt{2})]$, $S = \text{diag}(\sqrt{10}, 0)$, $V = [(1/\sqrt{5}, -2/\sqrt{5}), (2/\sqrt{5}, 1/\sqrt{5})]$

Metoda najmniejszych kwadratów, χ^2

Dane (x_i, y_i) , $i = 1, 2, \dots, N$ modelujemy za pomocą linii prostej $y(x) = a + bx$ minimalizując funkcję χ^2 :

$$\chi^2(a, b) = \sum_{i=1}^N \left(\frac{y_i - a - bx}{\sigma_i} \right)^2 \quad (10)$$

gdzie σ_i jest błędem wielkości y_i . Wartości parametrów a i b otrzymamy, minimalizując χ^2 . Wprowadzając oznaczenia

$$\begin{aligned} S &= \sum_{i=1}^N 1/\sigma_i^2, & S_x &= \sum x_i/\sigma_i^2, & S_y &= \sum y_i/\sigma_i^2, \\ S_{xx} &= \sum x_i^2/\sigma_i^2, & S_{xy} &= \sum x_i y_i/\sigma_i^2, \end{aligned} \quad (11)$$

otrzymamy

$$\begin{aligned} a &= \frac{S_{xx} S_y - S_x S_{xy}}{\Delta}, \\ b &= \frac{S S_{xy} - S_x S_y}{\Delta}, \end{aligned} \quad (12)$$

gdzie $\Delta = S S_{xx} - S_x^2$.

Błędy parametrów a i b i

Analiza propagacji błędów w danych y_i pokazuje, że błędy w dowolnej funkcji $f = f(x_i, y_i)$ określa wariancja σ_f^2

$$\sigma_f^2 = \sum_{i=1}^N \sigma_i^2 \left(\frac{\partial f}{\partial y_i} \right)^2. \quad (13)$$

W rozważanym przypadku

$$\sigma_a^2 = S_{xx}/\Delta, \quad \sigma_b^2 = S/\Delta. \quad (14)$$

Dodatkowo, tzw. kowariancja a i b jest

$$\text{Cov}(a, b) = -S_x/\Delta. \quad (15)$$

Współczynnik korelacji r_{ab} między nieokreślonością a i nieokreślonością b jest równy

$$r_{ab} = \frac{-S_x}{\sqrt{SS_{xx}}}. \quad (16)$$

$r_{ab} > 0$ – błędy w a i b są prawdopodobnie jednakowego znaku (*korelacja*)

$r_{ab} < 0$ – błędy są przeciwnego znaku (*antykorrelacja*).

Dobroć dopasowania modelu do danych (y_i) pozwala ocenić znaczenie parametrów. Prawdopodobieństwo, że wartość χ^2 (Równanie 10) jest przypadkowa jest dane przez

$$Q\left(\frac{N-2}{2}, \frac{\chi^2}{2}\right), \quad (17)$$

gdzie

$$Q(a, x) = \frac{\Gamma(a, x)}{\Gamma(a)}. \quad (18)$$

Tutaj $\Gamma(\dots)$ jest funkcją gamma Eulera lub niepełną funkcją gamma:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (19)$$

$$\Gamma(a, x) = \int_0^x e^{-t} t^{a-1} dt. \quad (20)$$

Metoda najmniejszych kwadratów, program

Formuły (12) wnoszą do obliczeń błędy numeryczne. Zapiszemy je inaczej (do celów programistycznych), definiując

$$t_i = \frac{1}{\sigma_i} \left(x_i - \frac{S_x}{S} \right), \quad i = 1, 2, \dots, N, \quad S_{tt} = \sum_{i=1}^N t_i^2. \quad (21)$$

Wtedy

$$b = \frac{1}{S_{tt}} \sum \frac{t_i y_i}{\sigma_i}, \quad a = \frac{S_y - b S_x}{S} \quad (22)$$

$$\sigma_a^2 = \frac{1}{S} \left(1 + \frac{S_x^2}{S S_{tt}} \right), \quad \sigma_b^2 = \frac{1}{S_{tt}}, \quad (23)$$

$$\text{Cov}(a, b) = -\frac{S_x}{S S_{tt}}, \quad r_{ab} = \frac{\text{Cov}(a, b)}{\sigma_a \sigma_b} \quad (24)$$

Zadania.

1. Sprawdzić formuły (21)–(24).
2. Napisać program `fit(...)`, który oblicza wielkości a , b , σ_a^2 , σ_b^2 , r_{ab} , oraz $\text{Cov}(a, b)$ dla zbioru danych (x_i, y_i) , $i = 1, 2, \dots, N$.

Metoda najmniejszych kwadratów, SVD

Dane $(t_i, y_i), i = 1, 2, \dots, m$ modelujemy za pomocą funkcji

$$f(t) = c_1\phi_1(t) + c_2\phi_2(t) + \dots + \phi_n(t), \quad (25)$$

gdzie ϕ_j są zadanymi funkcjami, a c_i parametrami modelu. Oznaczmy

$$a_{ij} = \phi_j(t_i). \quad (26)$$

Jeśli y oznacza wektor o elementach y_i , a c wektor o składowych c_i to równanie

$$\sum_j c_j \phi_j(t_i) \approx y_i, \quad i = 1, 2, \dots, m, \quad (27)$$

przybliża dane (t_i, y_i) . Można je zapisać w postaci

$$Ac \approx y \quad (28)$$

Równanie to rozwiążemy metodą SVD macierzy A .