

Wstęp do metod numerycznych

Singular Value Decomposition

P. F. Góra

<http://th-www.if.uj.edu.pl/zfs/gora/>

2019

Osobliwe układy równań

Dany jest układ równań

$$\begin{bmatrix} \frac{7}{6} & -\frac{1}{3} & -\frac{5}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{5}{6} & -\frac{1}{3} & \frac{7}{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \quad (1)$$

Jak łatwo sprawdzić, wyznacznik główny tego układu równań wynosi zero, a więc układ równań (1) nie ma jednoznacznego rozwiązania. Ale może ma on rozwiązanie niejednoznaczne? Aby tak było, wszystkie jego wyznaczniki poboczne powinny zniknąć. Sprawdzamy:

$$\det \begin{bmatrix} 1 & -\frac{1}{3} & -\frac{5}{6} \\ 1 & \frac{2}{3} & -\frac{1}{3} \\ 1 & -\frac{1}{3} & \frac{7}{6} \end{bmatrix} = 2 \neq 0, \quad (2)$$

a więc układ równań (1) jest sprzeczny (nie ma rozwiązań). Z drugiej strony układ równań

$$\begin{bmatrix} \frac{7}{6} & -\frac{1}{3} & -\frac{5}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{5}{6} & -\frac{1}{3} & \frac{7}{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

ma rozwiązanie, ale jest ono niejednoznaczne (wszystkie wyznaczniki poboczne równają się zero).

Układy równań (1) i (3) różnią się tylko prawymi stronami: ich macierz jest taka sama (i jest osobliwa). Widzimy, że w przypadku osobliwej macierzy układu równań to prawa strona decyduje, czy układ ma rozwiązania, czy też ich nie ma; *domyślamy się*, że to, jakie prawe strony oznaczają istnienie rozwiązań, jakie zaś ich brak, musi jakoś być związane z własnościami (osobliwej) macierzy układu równań. *Wygodnie byłoby mieć metodę pozwalającą na **zidentyfikowanie** wektorów, dla których osobliwy układ równań ma rozwiązania* (wzory Cramera pozwalają jedynie *sprawdzić*, czy dana prawa strona prowadzi do istnienia rozwiązań). *Metodą taką jest **SVD**.*

Singular Value Decomposition

Twierdzenie 1. *Dla każdej macierzy $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M \geq N$, istnieje faktoryzacja*

$$\mathbf{A} = \mathbf{U} [\text{diag}(w_i)] \mathbf{V}^T, \quad (4)$$

gdzie $\mathbf{U} \in \mathbb{R}^{M \times N}$ jest macierzą kolumnowo ortogonalną, $\mathbf{V} \in \mathbb{R}^{N \times N}$ jest macierzą ortogonalną oraz $w_i \in \mathbb{R}$, $i = 1, \dots, N$. Rozkład ten nazywamy rozkładem względem wartości osobliwych (Singular Value Decomposition, SVD). Jeżeli $M = N$, macierz \mathbf{U} jest macierzą ortogonalną.

Nie przedstawiamy tu algorytmu tej faktoryzacji — w razie potrzeby można skorzystać z którejś ze sprawdzonych procedur bibliotecznych.

Złożoność obliczeniowa faktoryzacji *SVD* wynosi $O(N^3)$, ale współczynnik przy wyrazie wiodącym jest wysoki, więc nie jest to metoda “z wyboru” do rozwiązywania układów równań. Znajduje ona jednak ważne zastosowania w przypadku osobliwych lub źle uwarunkowanych układów równań. Aby to zrozumieć, trzeba najpierw omówić algebraiczne własności tej faktoryzacji.

Jądro i zasięg operatora

Niech $\mathbf{A} \in \mathbb{R}^{M \times N}$. *Jądrem operatora \mathbf{A}* nazywam

$$\text{Ker } \mathbf{A} = \{\mathbf{x} \in \mathbb{R}^N : \mathbf{A}\mathbf{x} = \mathbf{0}\}. \quad (5)$$

Zasięgiem operatora \mathbf{A} nazywam

$$\text{Range } \mathbf{A} = \{\mathbf{y} \in \mathbb{R}^M : \exists \mathbf{x} \in \mathbb{R}^N : \mathbf{A}\mathbf{x} = \mathbf{y}\}. \quad (6)$$

Jądro i zasięg operatora są przestrzeniami liniowymi. Jeśli $M = N < \infty$,
 $\dim(\text{Ker } \mathbf{A}) + \dim(\text{Range } \mathbf{A}) = N$.

Sens SVD

Sens *SVD* najlepiej widać w przypadku, w którym co najmniej jedna z wartości $w_i = 0$. Aby to zrozumieć, zobaczmy jak macierz $\mathbf{A} \in \mathbb{R}^{M \times N}$, posiadająca faktoryzację (4), działa na pewien wektor $\mathbf{x} \in \mathbb{R}^N$:

$$\mathbf{Ax} = \mathbf{U} [\text{diag}(w_i)] \mathbf{V}^T \mathbf{x} \quad (7)$$

Dla ustalenia uwagi, niech $w_1 = 0$, $w_{i>1} \neq 0$.

Przypomnienie z algebry ☹

Macierz $\mathbf{V} \in \mathbb{R}^{N \times N}$ jest macierzą *ortogonalną*, co oznacza, że jej kolejne kolumny stanowią ortonormalną bazę w \mathbb{R}^N . Wektor

$$\mathbf{z} = \mathbf{V}^T \mathbf{x} \quad (8)$$

stanowi rozkład wektora \mathbf{x} w bazie kolumn macierzy \mathbf{V} . Wyrażenie (7) możemy przepisać jako

$$\mathbf{Ax} = \mathbf{U} [\text{diag}(w_i)] \mathbf{z}, \quad (9)$$

gdzie \mathbf{z} spełnia (8).

Sens SVD — przypadek 1

Przypuśćmy, że wektor \mathbf{x} jest równy pierwszej kolumnie macierzy \mathbf{V} . Wówczas $\mathbf{z} = [1, 0, 0, \dots, 0]^T$, gdyż \mathbf{x} jest prostopadły do wszystkich kolumn \mathbf{V} poczynając od drugiej. W takim wypadku

$$\mathbf{Ax} = \mathbf{U} \begin{bmatrix} 0 & & & & & \\ & w_2 & & & & \\ & & w_3 & & & \\ & & & \ddots & & \\ & & & & w_N & \\ & & & & & & \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (10)$$

Zero z lewego górnego rogu macierzy diagonalnej zabije jedynekę w wektorze po prawej stronie tej równości, a z kolei zera z tego wektora zabiją niezerowe w_i . Ostatecznie $\mathbf{Ax} = \vec{0}$. Stało się tak dlatego, że wektor \mathbf{x} był równy kolumnie macierzy \mathbf{V} , odpowiadającej zerowemu współczynnikowi w_i . *Kolumny macierzy \mathbf{V} , odpowiadające zerowym współczynnikom w_i , stanowią bazę w jądrze operatora \mathbf{A} .*

Sens SVD — przypadek 2

Niech teraz wektor \mathbf{x} będzie “dowolny”, to znaczy nie pozostaje w jakiejś specjalnej relacji do macierzy \mathbf{V} . Wówczas $\mathbf{z} = \mathbf{V}^T \mathbf{x} = [z_1, z_2, z_3, \dots, z_N]^T$, a zatem z (9) mamy

$$\mathbf{Ax} = \mathbf{U} \begin{bmatrix} 0 & & & & & \\ & w_2 & & & & \\ & & w_3 & & & \\ & & & \dots & & \\ & & & & w_N & \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{bmatrix} = \mathbf{U} \begin{bmatrix} 0 \\ w_2 z_2 \\ w_3 z_3 \\ \vdots \\ w_N z_N \end{bmatrix}. \quad (11)$$

Wynikiem ostatniego mnożenia będzie pewien wektor z przestrzeni \mathbb{R}^M . Ponieważ pierwszym elementem wektora $[0, w_2 z_2, \dots, w_N z_N]^T$ jest zero, **wynik ten nie zależy od pierwszej kolumny macierzy \mathbf{U}** . Kolumny macierzy \mathbf{U} są wzajemnie ortogonalne, a wektor \mathbf{Ax} nie zawiera przyczynku wzdłuż wektora będącego pierwszą kolumną \mathbf{U} , a jedynie przyczynki od wektorów

prostopadłych do tej kolumny. Widzimy zatem, że *kolumny macierzy U , odpowiadające niezerowym współczynnikom w_i , stanowią bazę w zasięgu operatora A .*

Dla macierzy osobliwych sens SVD sprowadza się do konstrukcji baz w jądrze i w zasięgu takiej macierzy. Jak za chwilę zobaczymy, można to uogólnić na przypadek macierzy nieosobliwych, ale bardzo źle uwarunkowanych.

Przykład

Macierz układów równań (1) i (3) ma następujący rozkład względem wartości osobliwych:

$$\begin{bmatrix} \frac{7}{6} & -\frac{1}{3} & -\frac{5}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{5}{6} & -\frac{1}{3} & \frac{7}{6} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (12)$$

Widzimy, że wektor $[1, 1, 1]^T$ rozpiną jądro tej macierzy, natomiast kombinacje liniowe wektorów $[-1, 2, -1]^T$, $[-1, 0, 1]^T$ stanowią jej zasięg.

Ponieważ rozważana macierz jest symetryczna i rzeczywista (i nieujemnie określona), jej rozkład *SVD* jest zarazem jej rozkładem diagonalnym (w przypadku ogólnym taka zgodność nie zachodzi).

SVD i odwrotność macierzy

Niech $\mathbf{A} \in \mathbb{R}^{N \times N}$. Zauważmy, że $|\det \mathbf{A}| = \prod_{i=1}^N w_i$, a zatem $\det \mathbf{A} = 0$ wtedy i tylko wtedy, gdy co najmniej jeden $w_i = 0$. Niech $\det \mathbf{A} \neq 0$. Wówczas równanie $\mathbf{A}\mathbf{x} = \mathbf{b}$ ma rozwiązanie postaci

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V} [\text{diag}(w_i^{-1})] \mathbf{U}^T \mathbf{b}. \quad (13)$$

Niech teraz $\det \mathbf{A} = 0$. Równanie $\mathbf{A}\mathbf{x} = \mathbf{b}$ *także* ma rozwiązanie, o ile tylko $\mathbf{b} \in \text{Range } \mathbf{A}$. Rozwiązanie to ma postać $\mathbf{x} = \tilde{\mathbf{A}}^{-1}\mathbf{b}$, gdzie

$$\tilde{\mathbf{A}}^{-1} = \mathbf{V} [\text{diag}(\tilde{w}_i^{-1})] \mathbf{U}^T. \quad (14a)$$

gdzie

$$\tilde{w}_i^{-1} = \begin{cases} w_i^{-1} & \text{gdy } w_i \neq 0, \\ 0 & \text{gdy } w_i = 0. \end{cases} \quad (14b)$$

SVD i macierze osobliwe

Wróćmy jeszcze raz do problemu osobliwych (z zerowym wyznacznikiem głównym) układów równań, wspomnianego już na stronie 14. Jeżeli $\det \mathbf{A} = 0$, układ równań z całą pewnością nie ma *jednoznacznego* rozwiązania. Może jednak mieć rozwiązanie (a nawet nieskończenie wiele rozwiązań), jeżeli prawa strona *należy do zasięgu* \mathbf{A} . Jest to równoważne warunkowi, że wszystkie wyznaczniki poboczne we wzorach Cramera zerują się. Wówczas **rozwiązaniem** układu równań jest każdy wektor postaci

$$\mathbf{x} = \tilde{\mathbf{A}}^{-1} \mathbf{b} + \mathbf{x}_0, \quad (15)$$

gdzie $\tilde{\mathbf{A}}^{-1}$ jest pseudoodwrotnością daną przez (14), zaś $\mathbf{x}_0 \in \text{Ker} \mathbf{A}$ jest dowolnym wektorem należącym do jądra. Rozwiązanie z $\mathbf{x}_0 = 0$ ma spośród nich najmniejszą normę. Zauważmy, że na wektory należące do zasięgu, pseudoodwrotność działa jak zwykła odwrotność macierzy.

Jeżeli b *nie* należy do zasięgu, wyrażenie (15) z $x_0 = 0$ daje rozwiązanie przybliżone i najlepsze w sensie najmniejszych kwadratów, co niekiedy jest bardzo użyteczne.

Przykład

Zastosowanie pseudoodwrotności (14) do układu (3) daje

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \quad (16)$$

W tym wypadku pseudoodwrotność zachowuje się jak odwrotność (prawa strona równania (3) należy do zasięgu) i rozwiązanie (16) jest ścisłe.

Do rozwiązania (16) można dodać dowolny wektor postaci $[\alpha, \alpha, \alpha]$, a więc należący do jądra.

SVD i współczynnik uwarunkowania

Twierdzenie 2. Jeżeli macierz $\mathbf{A} \in \mathbb{R}^{N \times N}$ posiada rozkład (4) oraz $\det \mathbf{A} \neq 0$, jej współczynnik uwarunkowania spełnia

$$\kappa = \frac{\max_i |w_i|}{\min_i |w_i|}. \quad (17)$$

Jeśli macierz jest źle uwarunkowana, ale *formalnie* odwracalna, numeryczne rozwiązanie równania $\mathbf{A}\mathbf{x} = \mathbf{b}$ może być zdominowane przez wzmocniony błąd zaokrąglenia. Aby tego uniknąć, często zamiast (bezużytecznego!) rozwiązania dokładnego (13), używa się *przybliżonego* (i użytecznego!) rozwiązania w postaci (14) z następującą modyfikacją

$$\tilde{w}_i^{-1} = \begin{cases} w_i^{-1} & \text{gdy } |w_i| > \tau, \\ 0 & \text{gdy } |w_i| \leq \tau, \end{cases} \quad (18)$$

gdzie τ jest pewną zadaną tolerancją.

Przykład

Mamy rozwiązać następujące dwa układy równań:

$$\begin{bmatrix} 0.666666667 & -0.166666666 & -0.333333333 \\ -0.166666666 & 0.166666667 & -0.166666666 \\ -0.333333333 & -0.166666666 & 0.666666667 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.284457048 \\ -0.577350273 \\ -0.129756514 \end{bmatrix} \quad (19a)$$

$$\begin{bmatrix} 0.666666667 & -0.166666666 & -0.333333333 \\ -0.166666666 & 0.166666667 & -0.166666666 \\ -0.333333333 & -0.166666666 & 0.666666667 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1.284457052 \\ -0.577350265 \\ -0.129756510 \end{bmatrix} \quad (19b)$$

Równania te różnią się tylko prawymi stronami, przy czym norma różnicy prawych stron jest rzędu 10^{-8} . Błąd takich rozmiarów łatwo może pojawić się w wyniku jakichś poprzednich obliczeń lub na skutek niepewności danych “zewnętrznych”, z którymi pracujemy.

Macierz w układach równań (19) jest symetryczna i dodatnio określona, a jej czynnik Cholesky’ego wynosi

$$\begin{bmatrix} 0.816496581 & & \\ -0.204124145 & 0.353553392 & \\ -0.408248290 & -0.707106777 & 0.000077460 \end{bmatrix} \quad (20)$$

Rozwiązania równań (19) za pomocą faktoryzacji Cholesky'ego mają postać

$$\mathbf{x}_a = \begin{bmatrix} -0.179434106 \\ -5.237183389 \\ -1.593647668 \end{bmatrix}, \quad \mathbf{x}_b = \begin{bmatrix} 3.903048668 \\ 2.927782158 \\ 2.488835105 \end{bmatrix}. \quad (21)$$

Różnica rozwiązań jest, co do normy, rzędu 10^{-10} , czyli jest rzędu 10^9 razy większa, niż różnica prawych stron.

Faktoryzacja *SVD* macierzy z układów (19) pokazuje, że wartości szczególne tej macierzy są w przybliżeniu równe $1, \frac{1}{2}, 10^{-9}$. Jeśli do rozwiązania układów równań (19) zastosować pseudoodwrotność (18) ($\tau = 10^{-8}$), w obu wypadkach otrzymamy

$$\mathbf{x} = \begin{bmatrix} 1.861807320 \\ -1.154700538 \\ 0.447593757 \end{bmatrix}. \quad (22)$$

(22) jest jedynie *przybliżonym* rozwiązaniem równań (19). Jest ono jednak bardziej użyteczne, niż “ściśle” rozwiązania (21). Te dwa ostatnie najwyraźniej są zdominowane przez błąd, jaki wystąpił wzdłuż kierunku odpowiadającego najmniejszej wartości szczególnej macierzy. Nie wiemy — i nie mamy sposobu, aby to stwierdzić — które z dwu rozwiązań (21) jest “poprawne”. Przybliżone rozwiązanie (22) po prostu ignoruje wpływ tego kierunku, a więc i zaburzeń wzdłuż niego występujących.

Zauważmy, że jeśli podstawimy rozwiązanie (22) do lewych stron równań (19), w obu wypadkach otrzymamy

$$\begin{bmatrix} 1.2844571 \\ -0.5773503 \\ -0.12975654 \end{bmatrix} \quad (23)$$

Norma różnicy tego wektora i *obu* prawych stron równań (19) wynosi w przybliżeniu $6.664 \cdot 10^{-8}$.

Nadokreślone układy równań

Niech $\mathbf{A} \in \mathbb{R}^{M \times N}$, $M > N$, $\mathbf{b} \in \mathbb{R}^M$, $\mathbf{x} \in \mathbb{R}^N$. Wówczas układ równań

$$\mathbf{Ax} = \mathbf{b} \quad (24)$$

ma więcej równań, niż niewiadomych. Układ taki, zwany nadokreślonym, w ogólności nie ma rozwiązań. Za pomocą SVD można jednak znaleźć jego rozwiązanie przybliżone. Mianowicie

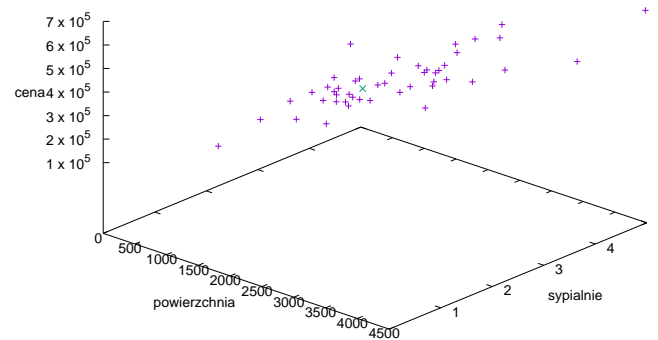
$$\|\mathbf{A} (\tilde{\mathbf{A}}^{-1} \mathbf{b}) - \mathbf{b}\|_2 = \text{minimum}, \quad (25)$$

gdzie $\tilde{\mathbf{A}}^{-1}$ jest pseudoodwrotnością (14). Widzimy, że $\tilde{\mathbf{A}}^{-1} \mathbf{b}$ jest przybliżonym, najlepszym w sensie najmniejszych kwadratów rozwiązaniem układu (24). Metoda ta jest *powszechnie* używana w liniowym zagadnieniu najmniejszych kwadratów.

Przykład

Ceny domów w Portland*.

W pliku znajdują się dane dotyczące cen domów. Cena zależy od powierzchni użytkowej domu i od liczby sypialni. Mamy do dyspozycji 47 danych. Chcemy się dowiedzieć, ile *powinna* wynosić cena domu o powierzchni 1650 sq. ft, z trzema sypialniami. Rysunek przedstawia dostępne dane:



*Przykład pochodzi z kursu Machine Learning na Uniwersytecie Stanforda. Dane są autentyczne.

Widzimy, że dane *w przybliżeniu* układają się na pewnej płaszczyźnie. Proponujemy wobec tego model:

$$\text{cena} = \theta_0 + \text{powierzchnia} \cdot \theta_1 + \text{sypialnie} \cdot \theta_2 \quad (26)$$

Równania tego typu powinny być spełnione dla *wszystkich* dostępnych danych:

$$\theta_0 + p_1\theta_1 + s_1\theta_2 = c_1 \quad (27a)$$

$$\theta_0 + p_2\theta_1 + s_2\theta_2 = c_2 \quad (27b)$$

⋮

$$\theta_0 + p_{47}\theta_1 + s_{47}\theta_2 = c_{47} \quad (27c)$$

Jest to układ 47 równań na trzy niewiadome $(\theta_0, \theta_1, \theta_2)$ i jest niezwykle mało prawdopodobne, że uda się go rozwiązać ściśle, to znaczy tak, aby

każde z równań (27a) było ściśle spełnione. Musimy zadowolić się rozwiązaniem przybliżonym.

W postaci stabelaryzowanej dane mają postać

Nr	powierznia	sypialnie	cena
1	2100	3	399900
2	1600	3	329900
3	2400	3	369000
.....			
45	852	2	179900
46	1852	4	299900
47	1203	3	239500

Uwaga: Wśród danych wejściowych, jedna (powierzchnia) przybiera wartości rzędu tysięcy, druga (sypialnie) rzędu jednośc. Może to, potencjalnie, być źródłem błędów numerycznych. Dlatego dane warto *znormalizować*, to znaczy odjąć od nich wartości średnie poszczególnych wielkości i podzielić przez odchylenia standardowe.

Równania (27a) zapisujemy w postaci macierzowej:

$$\mathbf{X} \cdot \boldsymbol{\theta} = \mathbf{c}, \quad (28)$$

gdzie

$$\mathbf{X} = \begin{bmatrix} 1 & p_1 & s_1 \\ 1 & p_2 & s_2 \\ 1 & p_3 & s_3 \\ \dots & \dots & \dots \\ 1 & p_{45} & s_{45} \\ 1 & p_{46} & s_{46} \\ 1 & p_{47} & s_{47} \end{bmatrix}$$

$\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2]^T$, $\mathbf{c} = [c_1, c_2, c_3, \dots, c_{45}, c_{46}, c_{47}]^T$, a p_i, s_i, c_i są (znormalizowaną) powierzchnią, liczbą sypialni i ceną i -tego domu. \mathbf{X} jest macierzą o 47 wierszach i 3 kolumnach; pierwsza kolumna macierzy \mathbf{X} zawiera same jedynki.

Korzystając z naszego ulubionego pakietu obliczającego faktoryzację *SVD* macierzy \mathbf{X} , do równania (28) stosujemy pseudoodwrotność (14) i znajdujemy *optymalne* rozwiązanie przybliżone

$$\boldsymbol{\theta}_{\text{opt}} = \tilde{\mathbf{X}}^{-1} \mathbf{c} \quad (29)$$

Następnie podstawiamy (znormalizowane) dane wejściowe ($p = 1650$, $s = 3$) i (po denormalizacji) uzyskujemy odpowiedź, iż taki dom powinien kosztować $\boldsymbol{\theta}_{\text{opt}} \cdot [1, p, s]^T = \$ 293081.46$.