

Matematyczne Metody FIZYKI I

Wykład 14

Zastosowania układów równań liniowych

Problem: Niech zbiór danych będzie reprezentowany przez n punktów na płaszczyźnie:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Chcemy znaleźć wielomian stopnia $n-1$: $p(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$

który przechodzi przez te punkty. Jeśli wszystkie współrzędne x są różne, wtedy istnieje dokładnie jeden wielomian stopnia $n-1$ lub niższego przechodzący przez te punkty.

Jego współczynniki wyznaczamy z układu równań:

$$\begin{cases} a_0 + a_1x_1 + a_2x_1^2 + \dots + a_{n-1}x_1^{n-1} = y_1 \\ a_0 + a_1x_2 + a_2x_2^2 + \dots + a_{n-1}x_2^{n-1} = y_2 \\ \vdots \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_{n-1}x_n^{n-1} = y_n \end{cases}$$

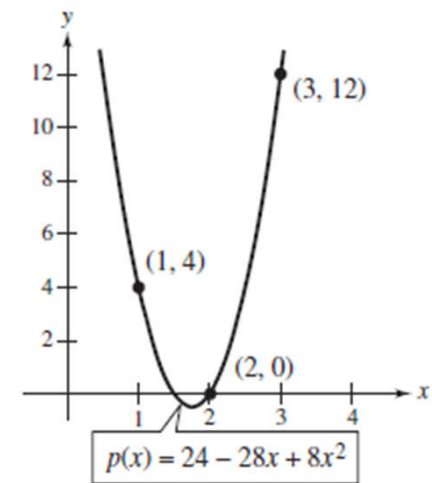
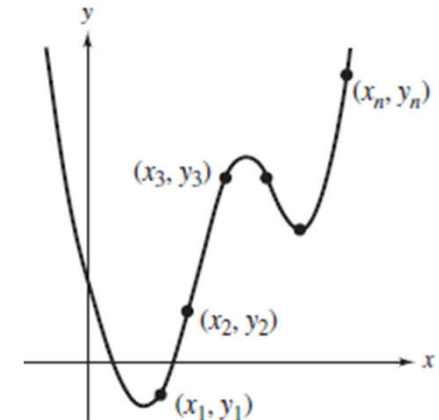
Przykład: Znajdź wielomian stopnia drugiego przechodzący przez punkty: $(1,4)$, $(2,0)$, $(3,12)$.

$$p(1) = a_0 + a_1 + a_2 = 4$$

$$p(2) = a_0 + 2a_1 + 4a_2 = 0$$

$$p(3) = a_0 + 3a_1 + 9a_2 = 12$$

Rozwiązując układ równań znajdujemy współczynniki :



Dopasowanie linii prostej

Problem: Znajdź linię prostą, która jest najlepiej dopasowana do punktów:

$$(1,1), (2,2), (3,4), (4,4), (5,6)$$

Jako miarę dopasowania linii prostej do punktów wyznaczamy sumę kwadratów odstępstw wartości funkcji o punktów

doświadczalnych:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - f(x_i))^2$$

Znajdujemy odpowiednie błędy dla powyższych dwóch prostych:

$$f(x) = 0.5 + x \Rightarrow \sum_{i=1}^5 (y_i - f(x_i))^2 = 1.25$$

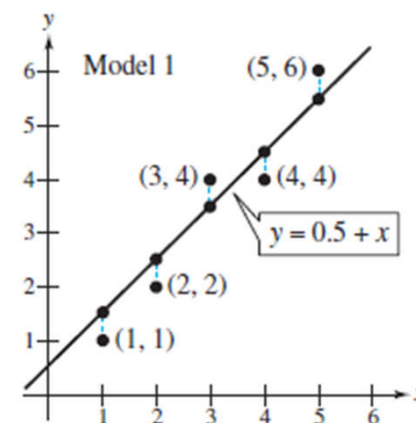
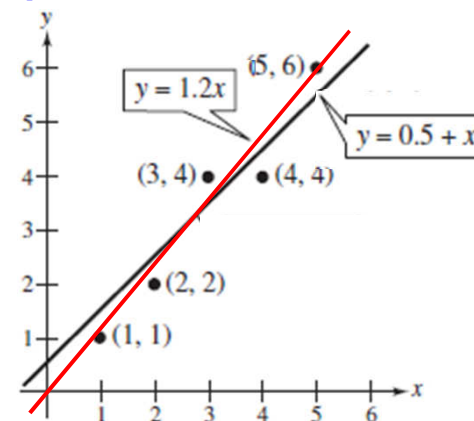
$$f(x) = 1.2x \Rightarrow \sum_{i=1}^5 (y_i - f(x_i))^2 = 1.00$$

Definicja: Prosta regresji metody najmniejszych kwadratów dla zbioru punktów (x_i, y_i) , $i = 1, \dots, n$ nazywamy funkcję

$$f(x) = a_0 + a_1 x$$

która minimalizuje sumę kwadratów błędów

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - f(x_i))^2$$



Metoda najmniejszych kwadratów

Ogólnie możemy napisać:

$$\begin{aligned}y_1 &= f(x_1) + [y_1 - f(x_1)] = a_0 + a_1 x_1 + e_1 \\y_2 &= f(x_2) + [y_2 - f(x_2)] = a_0 + a_1 x_2 + e_2 \\&\vdots \\y_n &= f(x_n) + [y_n - f(x_n)] = a_0 + a_1 x_n + e_n\end{aligned}$$

lub w postaci macierzowej $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}$ gdzie

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

Twierdzenie: Dla modelu regresji liniowej $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E}$ współczynniki linii regresji metody najmniejszych kwadratów dane są przez

$$\mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y}$$

natomiast suma kwadratów błędów przez $\mathbf{E}^T \mathbf{E}$.

Metoda najmniejszych kwadratów

Przykład: Znajdź prostą regresji metody najmniejszych kwadratów dla punktów
(1,1), (2,2), (3,4), (4,4), (5,6)

$$\mathbf{Y} = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 4 \\ 6 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} = \begin{pmatrix} -0.2 \\ 1.2 \end{pmatrix}$$

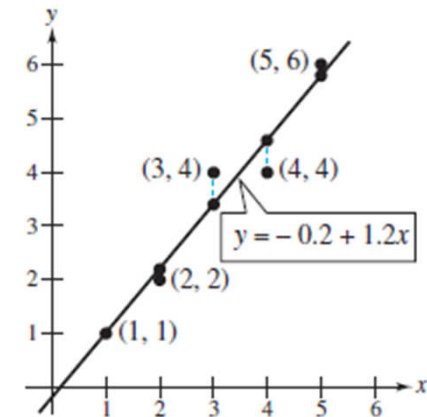
A więc prosta regresji metody najmniejszych kwadratów ma postać:

$$y = -0.2 + 1.2x$$

Natomiast suma kwadratów błędów wynosi:

$$\begin{aligned} \mathbf{E}^T \mathbf{E} &= (\mathbf{Y} - \mathbf{A} \mathbf{X})^T (\mathbf{Y} - \mathbf{A} \mathbf{X}) = \\ &= \left(\mathbf{Y} - \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \right)^T \left(\mathbf{Y} - \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{Y} \right) = 0.8 \end{aligned}$$

I jest mniejsza od wyznaczonych wcześniej dla innych prostych.



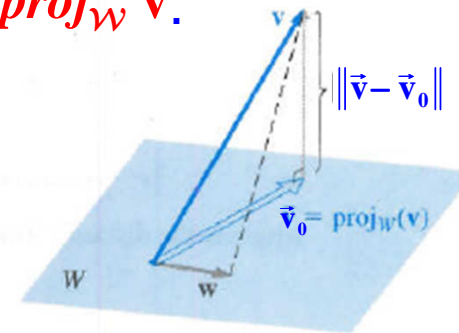
Najlepsze przybliżenie

Definicja: Jeśli \mathcal{W} jest podprzestrzenią przestrzeni liniowej \mathcal{V} oraz $\vec{v} \in \mathcal{V}$ to **najlepszym przybliżeniem** wektora \vec{v} w \mathcal{W} jest wektor \vec{v}_0 taki, że dla każdego wektora $\vec{w} \in \mathcal{W}$ różnego od \vec{v}_0 zachodzi $\|\vec{v} - \vec{v}_0\| < \|\vec{v} - \vec{w}\|$

Twierdzenie: Jeśli \mathcal{W} jest podprzestrzenią przestrzeni \mathcal{V} z iloczynem skalarnym oraz $\vec{v} \in \mathcal{V}$ wówczas najlepszym przybliżeniem wektora \vec{v} w \mathcal{W} jest $\mathit{proj}_{\mathcal{W}} \vec{v}$.

Dowód: Niech $\vec{w} \in \mathcal{W}$ oraz $\vec{w} \neq \mathit{proj}_{\mathcal{W}} \vec{v}$

$$\vec{v} - \vec{w} = \underbrace{(\vec{v} - \mathit{proj}_{\mathcal{W}} \vec{v})}_{\in \mathcal{W}^\perp} + \underbrace{(\mathit{proj}_{\mathcal{W}} \vec{v} - \vec{w})}_{\in \mathcal{W}}$$



Korzystając z twierdzenia Pitagorasa mamy:

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v} - \mathit{proj}_{\mathcal{W}} \vec{v}\|^2 + \|\mathit{proj}_{\mathcal{W}} \vec{v} - \vec{w}\|^2 \Rightarrow \|\vec{v} - \mathit{proj}_{\mathcal{W}} \vec{v}\| < \|\vec{v} - \vec{w}\|$$

Przykład: Dane są wektory $\vec{u}_1 = (1, 2, -1)^T$, $\vec{u}_2 = (5, -2, 1)^T$, $\vec{v} = (3, 2, 5)^T$

Znajdź najlepsze przybliżenie wektora \vec{v} w przestrzeni $\mathcal{W} = \mathit{span}(\vec{u}_1, \vec{u}_2)$

$$\mathit{proj}_{\mathcal{W}} \vec{v} = \left(\frac{\vec{u}_1 \cdot \vec{v}}{\vec{u}_1 \cdot \vec{u}_1} \right) \vec{u}_1 + \left(\frac{\vec{u}_2 \cdot \vec{v}}{\vec{u}_2 \cdot \vec{u}_2} \right) \vec{u}_2 = \frac{1}{5} \begin{pmatrix} 15 \\ -2 \\ 1 \end{pmatrix} \quad \|\mathit{perp}_{\mathcal{W}} \vec{v}\| = \|\vec{v} - \mathit{proj}_{\mathcal{W}} \vec{v}\| = \frac{12}{\sqrt{5}}$$

Metoda najmniejszych kwadratów

Rozważamy układ równań $A\vec{x} = \vec{b}$ gdzie $A_{m \times n}$ i $\vec{b} \in \mathcal{R}^m$

Ponieważ układ jest sprzeczny, więc można założyć, że $\vec{b} \notin \mathcal{S} = \mathcal{R}(A)$

Szukamy wektora \vec{x} który minimalizuje wielkość $\|A\vec{x} - \vec{b}\|$

Chcemy znaleźć taki wektor $A\vec{x} \in \mathcal{S}$ który jest najbliższy wektorowi \vec{b} .

Wiemy, że takim wektorem jest $A\vec{x}_0 = \text{proj}_{\mathcal{S}} \vec{b}$

Ponieważ wektor $A\vec{x}_0 - \vec{b} = \text{proj}_{\mathcal{S}} \vec{b} - \vec{b}$ jest ortogonalny do \mathcal{S} , więc

$$A\vec{x}_0 - \vec{b} \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

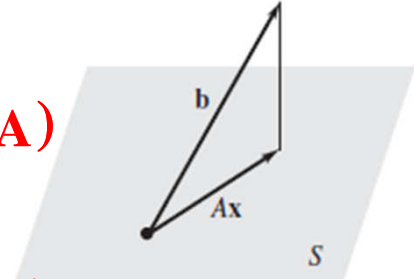
A to oznacza, że:

$$A^T(A\vec{x}_0 - \vec{b}) = 0 \Rightarrow A^T A\vec{x}_0 - A^T \vec{b} = 0 \Rightarrow \vec{x}_0 = (A^T A)^{-1} A^T \vec{b}$$

Metoda najmniejszych kwadratów sprowadza się do rozwiązania układu $n \times n$ równań liniowych $A^T A \vec{x} = A^T \vec{b}$. Te równania nazywane są **równaniami normalnymi** problemu najmniejszych kwadratów $A\vec{x} = \vec{b}$.

Układ równań normalnych nigdy nie jest sprzeczny, ale może być nieoznaczony.

Układ ten jest oznaczony gdy rząd macierzy $A_{m \times n}$ wynosi n .



Metoda najmniejszych kwadratów

Przykład: Korzystając z metody najmniejszych kwadratów dopasuj wielomian kwadratowy $y = a_0 + a_1x + a_2x^2$ do następujących danych:

rok	1980	1985	1990	1995	2000	2005
populacja	4.5	4.8	5.3	5.7	6.1	6.5

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \\ 1 & 20 & 400 \\ 1 & 25 & 625 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 4.5 \\ 4.8 \\ 5.3 \\ 5.7 \\ 6.1 \\ 6.5 \end{pmatrix}$$

$$A^T A \vec{x} = A^T \vec{b} \Rightarrow \begin{pmatrix} 6 & 75 & 1375 \\ 75 & 1375 & 28125 \\ 1375 & 28125 & 611875 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 32.9 \\ 447 \\ 8435 \end{pmatrix} \Rightarrow \vec{x} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 4.5 \\ 0.08 \\ 0 \end{pmatrix}$$

A więc metodą najmniejszych kwadratów otrzymujemy zależność liniową $y = 4.5 + 0.08x$

Przewidywanie dla roku 2010: $y = 4.5 + 0.08 \cdot 30 \approx 6.9$

Wartości osobliwe macierzy

Z dowolnej macierzy rzeczywistej $A_{m \times n}$ można utworzyć macierz symetryczną $A^T A$, a więc diagonalizowalną za pomocą ortogonalnej transformacji podobieństwa.

Wartości własne macierzy $A^T A$ są rzeczywiste i nieujemne:

$$0 \leq \|A \vec{v}\|^2 = (A \vec{v}) \cdot (A \vec{v}) = (A \vec{v})^T (A \vec{v}) = \vec{v}^T A^T A \vec{v} = \vec{v}^T \lambda \vec{v} = \lambda \|\vec{v}\|^2 = \lambda$$

Definicja: **Wartościami osobliwymi** macierzy $A_{m \times n}$ nazywamy pierwiastki kwadratowe wartości własnych macierzy $A^T A$, i oznaczamy przez $\sigma_1, \sigma_2, \dots, \sigma_n$ gdzie $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$.

Twierdzenie: **Niech** $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ będą wartościami własnymi symetrycznej macierzy $A_{n \times n}$ stowarzyszonej z formą kwadratową $f(\vec{x}) = \vec{x}^T A \vec{x}$. Jeśli $\|\vec{x}\| = 1$ to wówczas:

(a) $\lambda_1 \geq f(\vec{x}) \geq \lambda_n$

(b) $\max f(\vec{x}) = \lambda_1$ i występuje dla \vec{x} będącego wektorem własnym do w. własnej λ_1 .

(c) $\min f(\vec{x}) = \lambda_n$ i występuje dla \vec{x} będącego wektorem własnym do w. własnej λ_n .

D: (a) $f(\vec{x}) = \vec{x}^T A \vec{x} = \vec{x}^T Q^T A Q \vec{x} = \vec{y}^T D \vec{y} \Rightarrow \|\vec{y}\| = \sqrt{\vec{y}^T \vec{y}} = \sqrt{\vec{x}^T Q^T Q \vec{x}} = \sqrt{\vec{x}^T \vec{x}} = 1$

$$f(\vec{x}) = \vec{x}^T A \vec{x} = \vec{y}^T D \vec{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2 \leq \lambda_1 (y_1^2 + y_2^2 + \dots + y_n^2) = \lambda_1$$

(b) $A \vec{q}_1 = \lambda_1 \vec{q}_1 \Rightarrow f(\vec{q}_1) = \vec{q}_1^T A \vec{q}_1 = \vec{q}_1^T \lambda_1 \vec{q}_1 = \lambda_1 (\vec{q}_1^T \vec{q}_1) = \lambda_1$

Rozkład według wartości osobliwych

Przykład: Znajdź wartości osobliwe macierzy \mathbf{A} .

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow \mathbf{A}^T \mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \Rightarrow \begin{cases} \sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \\ \sigma_2 = \sqrt{\lambda_2} = 1 \end{cases}$$

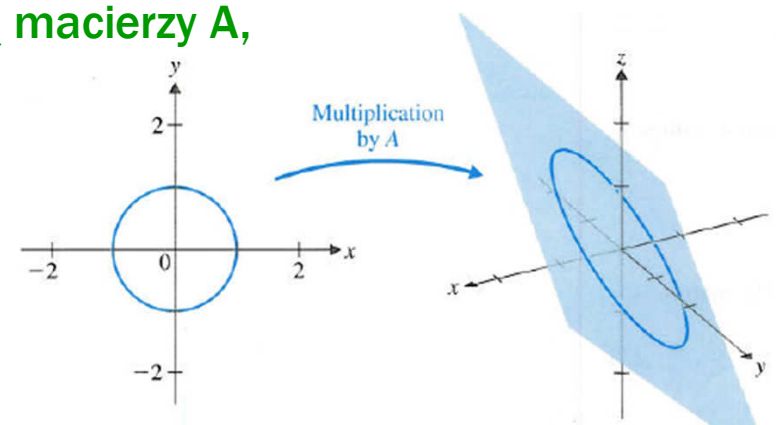
Niech $\|\vec{x}\| = 1$ wówczas: $\|\mathbf{A}\vec{x}\|^2 = (\mathbf{A}\vec{x})^T (\mathbf{A}\vec{x}) = \vec{x}^T \mathbf{A}^T \mathbf{A} \vec{x} = 2x_1^2 + 2x_1x_2 + 2x_2^2$

Maksimum i minimum tej formy kwadratowej wynosi odpowiednio $\lambda_1 = 3$ oraz $\lambda_2 = 1$ i występuje dla odpowiadających im wektorów własnych:

$$\vec{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ oraz } \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Transformacja liniowa dana określona za pomocą macierzy \mathbf{A} , przekształca \mathcal{R}^2 w płaszczyznę $x-y-z = 0$ w \mathcal{R}^3 .

Natomiast jednostkowy okrąg w \mathcal{R}^2 przechodzi w elipsę leżącą w płaszczyźnie $x-y-z = 0$ w \mathcal{R}^3 .



Rozkład według wartości osobliwych

Twierdzenie: Niech macierz $A_{m \times n}$ ma wartości osobliwe $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ oraz $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$. Wówczas istnieją ortogonalne macierze $U_{n \times n}$ i $V_{m \times m}$ oraz macierz $\Sigma_{m \times n}$ takie, że $A = U \Sigma V^T$

D: Konstruujemy macierz V z wektorów własnych macierzy $A^T A$: $V = [\vec{v}_1 \ \dots \ \vec{v}_n]$

Konstrukcja macierzy U :

$$(A \vec{v}_i) \cdot (A \vec{v}_j) = (A \vec{v}_i)^T (A \vec{v}_j) = \vec{v}_i^T A^T A \vec{v}_j = \vec{v}_i^T \lambda_j \vec{v}_j = \lambda_j (\vec{v}_i \cdot \vec{v}_j) = 0$$

$$\sigma_i = \|A \vec{v}_i\| \Rightarrow \vec{u}_i = \frac{1}{\sigma_i} A \vec{v}_i \quad \text{dla } i = 1, \dots, r$$

Zbiór wektorów \vec{u}_i uzupełniamy aby tworzył bazę ortonormalną w \mathcal{R}^m :

$$U = [\vec{u}_1 \ \dots \ \vec{u}_m]$$

$$AV = A[\vec{v}_1 \ \dots \ \vec{v}_n] = [A \vec{v}_1 \ \dots \ A \vec{v}_n] = [A \vec{v}_1 \ \dots \ A \vec{v}_r \ \mathbf{0} \ \dots \ \mathbf{0}] =$$

$$= [\sigma_1 \vec{u}_1 \ \dots \ \sigma_r \vec{u}_r \ \mathbf{0} \ \dots \ \mathbf{0}] = [\vec{u}_1 \ \dots \ \vec{u}_m] \begin{bmatrix} \sigma_1 & \dots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0} \\ \mathbf{0} & \dots & \sigma_r & \\ & & \mathbf{0} & \mathbf{0} \end{bmatrix} = U \Sigma$$

Rozkład według wartości osobliwych

Przykład: Znajdź rozkład macierzy A według wartości osobliwych.

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \Rightarrow \begin{cases} \sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \\ \sigma_2 = \sqrt{\lambda_2} = 1 \end{cases} \Rightarrow \vec{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Macierze Σ oraz V mają postać:

$$\Sigma = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Znajdujemy macierz U :

$$\vec{u}_1 = \frac{1}{\sigma_1} A \vec{v}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \quad \vec{u}_2 = \frac{1}{\sigma_2} A \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

Aby wektory \vec{u}_i tworzyły bazę ortonormalną w \mathcal{R}^3 znajdujemy trzeci wektor korzystając z metody Grama-Schmidta:

$$\vec{u}_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \Rightarrow U = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & 0 & -\sqrt{2} \\ 1 & -\sqrt{3} & \sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \end{pmatrix}$$

Łatwo sprawdzić, że $A = U \Sigma V^T$