Soal

Lakukan analisis PCA untuk data berikut, data berikut menggunakan dua fitur, x1 dan x2. Dengan menggunakan algoritma PCA, hitunglah proses terbentuknya reduksi dimensi dari N ke k yang baru serta tunjukkan nilai dimensi semula setelah direkonstruksi.

$$x_1 = [2{,}5 \quad 0{,}5 \quad 2{,}2 \quad 1{,}9 \quad 3{,}1 \quad 2{,}3 \quad 2{,}0 \quad 1{,}0 \quad 1{,}5 \quad 1{,}1]$$

$$x_2 = [2{,}4 \quad 0{,}7 \quad 2{,}9 \quad 2{,}2 \quad 3{,}0 \quad 2{,}7 \quad 1{,}6 \quad 1{,}1 \quad 1{,}6 \quad 0{,}9]$$

Jika ditulis dalam matriks:

Penyelesaian:

0. Tahapan persiapan data.
   Pertama untuk menjadikan kedua data tersebut dalam deretan dataset, secara horisontal membentuk baris data (baris = 10), secara vertical membentuk dua fitur (kolom = fitur = 2). Untuk itu dilakukan transpose terhadap data tersebut, diumpamakan bahwa himpunan datanya adalah X. Perubahan bentuk ini akan memudahkan transformasi perhitungan secara matriks

$$\begin{array}{cc} x_1 & x_2 \end{array}$$

$$X = \begin{bmatrix} 2{,}5 & 2{,}4 \\ 0{,}5 & 0{,}7 \\ 2{,}2 & 2{,}9 \\ 1{,}9 & 2{,}2 \\ 3{,}1 & 3{,}0 \\ 2{,}3 & 2{,}7 \\ 2{,}0 & 1{,}6 \\ 1{,}0 & 1{,}1 \\ 1{,}5 & 1{,}6 \\ 1{,}1 & 0{,}9 \end{bmatrix}_{10x2}$$

Algoritma PCA, sebagai berikut:

1. Menentukan mean global dari data.
   - Pencarian nilai rata-rata fitur 1 (x_1)

$$\bar{x}_1 = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}}{N}$$

$$\bar{x}_1 = \frac{x_{1.1} + x_{1.2} + x_{1.3} + x_{1.4} + x_{1.5} + x_{1.6} + x_{1.7} + x_{1.8} + x_{1.9} + x_{1.10}}{N}$$

$$\bar{x}_1 = \frac{2{,}5 + 0{,}5 + 2{,}1 + 1{,}9 + 3{,}1 + 2{,}3 + 2{,}0 + 1{,}0 + 1{,}5 + 1{,}1}{10} = \frac{18{,}1}{10} = 1{,}81$$

   - Pencarian nilai rata-rata fitur 2 (x_2)

$$\bar{x}_2 = \frac{x_{2.1} + x_{2.2} + x_{2.3} + x_{2.4} + x_{2.5} + x_{2.6} + x_{2.7} + x_{2.8} + x_{2.9} + x_{2.10}}{N}$$

$$\bar{x}_2 = \frac{2{,}4 + 0{,}7 + 2{,}9 + 2{,}2 + 3{,}0 + 2{,}7 + 1{,}6 + 1{,}1 + 1{,}6 + 0{,}9}{10} = \frac{19{,}1}{10} = 1{,}91$$

$$Mean\ Global\ Data = [1{,}81 \quad 1{,}91]$$

2. Menghitung Zero Mean (pengurangan antara data asli dengan nilai rata-rata).

$$Zero\ Mean = \begin{bmatrix} 2,5-1,81 & 2,4-1,91 \\ 0,5-1,81 & 0,7-1,91 \\ 2,2-1,81 & 2,9-1,91 \\ 1,9-1,81 & 2,2-1,91 \\ 3,1-1,81 & 3,0-1,91 \\ 2,3-1,81 & 2,7-1,91 \\ 2,0-1,81 & 1,6-1,91 \\ 1,0-1,81 & 1,1-1,91 \\ 1,5-1,81 & 1,6-1,91 \\ 1,1-1,81 & 0,9-1,91 \end{bmatrix}_{10x2} = \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix}_{10x2}$$

3. Membangun matriks covarians (perkalian antara zero mean dengan transposenya).

$$Covarians = \frac{1}{(N-1)} \emptyset^T \emptyset, \qquad \emptyset = Zero\ Mean$$

$$Covarians = \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix}^T_{10x2} * \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix}_{10x2}$$

$$= \begin{bmatrix} 2,5 & 0,5 & 2,2 & 1,9 & 3,1 & 2,3 & 2,0 & 1,0 & 1,5 & 1,1 \\ 2,4 & 0,7 & 2,9 & 2,2 & 3,0 & 2,7 & 1,6 & 1,1 & 1,6 & 0,9 \end{bmatrix}_{2x10} * \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix}_{10x2}$$

$$Covarians = C = \begin{bmatrix} 0,6165 & 0,6154 \\ 0,6154 & 0,7165 \end{bmatrix}$$

4. Menghitung Eigen Value (Nilai Eigen).

$$det[\ \lambda I - C] = 0$$

$$\lambda^2 - 1,333\lambda + 0,063005 = 0$$

$$\lambda_1 = \frac{1,333333 + 1,234855}{2} = 1,284094$$

$$\lambda_2 = \frac{1,333333 - 1,234855}{2} = 0,049239$$

$$Matriks\ Eigen\ Value = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1,284094 & 0 \\ 0 & 0,049239 \end{bmatrix}$$

5. Menghitung Eigen Vector.

$$CU = \lambda U$$

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \lambda \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \lambda \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} c_{11} - \lambda & c_{12} \\ c_{21} & c_{22} - \lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0$$

- Untuk Eigen Vector pertama dengan $\lambda_1 = 1,284094$; $C = \begin{bmatrix} 0,6165 & 0,6154 \\ 0,6154 & 0,7165 \end{bmatrix}$

$$(c_{11} - \lambda_1)u_1 + c_{12}u_2 = 0 \rightarrow (0,6165 - 1,284094)u_1 + 0,6154u_2 = 0$$

$$-0,66759u_1 + 0,6154u_2 = 0 \text{ ...(1)}$$

$$c_{21}u_1 + (c_{22} - \lambda_2)u_2 = 0 \rightarrow 0,6154u_1 + (0,7165 - 1,284094)u_2 = 0$$

$$0,6154u_1 - 0,56759u_2 = 0 \text{ ... (2)}$$

Persamaan 1 disubsstitusi ke persamaan 2.

$$-0,66759u_1 + 0,6154u_2 = 0 \text{ ...(1)}$$

$-0,66759u_1 = -0,6154u_2$; u_1 dan u_2 dibagi dengan 0,66759

ambil $u_2 = 1$

$$-0,66759u_1 = -0,6154u_2$$

$$0,66759u_1 = 0,6154(1)$$

$$u_1 = 0,921823; u_2 = 1$$

$$\sqrt{u_1{}^2 + u_2{}^2} = \sqrt{0,92182^2 + 1^2} = 1,360058$$

$$Vektor\ u_1 = \frac{0,921823}{1,360058} = 0,677782$$

$$Vektor\ u_2 = \frac{1}{1,360058} = 0,735263$$

Eigen vector dari data $\lambda_1 = 1,284094$ yaitu:

$$U = \begin{bmatrix} 0,677782 \\ 0,735263 \end{bmatrix}$$

- Untuk Eigen Vector kedua dengan $\lambda_2 = 0,049239$; $C = \begin{bmatrix} 0,6165 & 0,6154 \\ 0,6154 & 0,7165 \end{bmatrix}$

$$(c_{11} - \lambda_2)u_1 + c_{12}u_2 = 0 \rightarrow (0,6165 - 0,049239)u_1 + 0,6154u_2 = 0$$

$$0{,}567261u_1 + 0{,}6154u_2 = 0 \ldots(1)$$

$$c_{21}u_1 + (c_{22} - \lambda_2)u_2 = 0 \rightarrow 0{,}6154u_1 + (0{,}7165 - 0{,}049239)u_2 = 0$$

$$0{,}6154u_1 + 0{,}667261u_2 = 0 \ \ldots(2)$$

Persamaan 1 disubstitusi ke persamaan 2.

$$0{,}6154u_1 + 0{,}667261u_2 = 0 \text{ ; misal: } u_2 = 1$$

$$0{,}6154u_1 = -0{,}667261 * (1)$$

$$0{,}6154u_1 = -0{,}667261$$

$$u_1 = -1{,}084272$$

$$\sqrt{{u_1}^2 + {u_2}^2} = \sqrt{(-1{,}084272)^2 + 1^2} = 1{,}475007$$

$$Vektor\ u_1 = \frac{-1{,}084274}{1{,}475007} = -0{,}735096$$

$$Vektor\ u_2 = \frac{1}{1{,}475007} = 0{,}677963$$

Eigen vector dari data $\lambda_2 = 0{,}049239$ yaitu:

$$U = \begin{bmatrix} -0{,}735096 \\ 0{,}677963 \end{bmatrix}$$

Eigen vector global:

$$U = \begin{bmatrix} 0{,}677782 & -0{,}735096 \\ 0{,}735263 & 0{,}677963 \end{bmatrix}$$

6. Mentransformasi data asli ke dalam bentuk yang berbeda (data baru).

$$Z = Zero\ Mean\ * Global\ Eigen\ Vector$$

$$Z = \begin{bmatrix} 0{,}69 & 0{,}49 \\ -1{,}31 & -1{,}21 \\ 0{,}39 & 0{,}99 \\ 0{,}09 & 0{,}29 \\ 1{,}29 & 1{,}09 \\ 0{,}49 & 0{,}79 \\ 0{,}19 & -0{,}31 \\ -0{,}81 & -0{,}81 \\ -0{,}31 & -0{,}31 \\ -0{,}71 & -1{,}01 \end{bmatrix}_{10x2} * \begin{bmatrix} 0{,}677782 & -0{,}735096 \\ 0{,}735263 & 0{,}677963 \end{bmatrix} = \begin{bmatrix} 0{,}82794845 & -0{,}17501437 \\ -1{,}77756265 & 0{,}14264053 \\ 0{,}99224535 & 0{,}38449593 \\ 0{,}27422665 & 0{,}13045063 \\ 1{,}67577545 & -0{,}20929417 \\ 0{,}91297095 & 0{,}17539373 \\ -0{,}09915295 & -0{,}34983677 \\ -1{,}14456645 & 0{,}046277729 \\ -0{,}43804395 & 0{,}01771123 \\ -1{,}22384085 & -0{,}16282447 \end{bmatrix}$$

7. Merekonstruksi data baru ke data yang lama

$$Row\ Zero\ Mean\ Data = Z * U^T$$

$$Row\ Zero\ Mean\ Data = \begin{bmatrix} 0{,}82794845 & -0{,}17501437 \\ -1{,}77756265 & 0{,}14264053 \\ 0{,}99224535 & 0{,}38449593 \\ 0{,}27422665 & 0{,}13045063 \\ 1{,}67577545 & -0{,}20929417 \\ 0{,}91297095 & 0{,}17539373 \\ -0{,}09915295 & -0{,}34983677 \\ -1{,}14456645 & 0{,}046277729 \\ -0{,}43804395 & 0{,}01771123 \\ -1{,}22384085 & -0{,}16282447 \end{bmatrix} * \begin{bmatrix} 0{,}677782 & -0{,}735096 \\ 0{,}735263 & 0{,}677963 \end{bmatrix}^T$$

$$= \begin{bmatrix} 0{,}82794845 & -0{,}17501437 \\ -1{,}77756265 & 0{,}14264053 \\ 0{,}99224535 & 0{,}38449593 \\ 0{,}27422665 & 0{,}13045063 \\ 1{,}67577545 & -0{,}20929417 \\ 0{,}91297095 & 0{,}17539373 \\ -0{,}09915295 & -0{,}34983677 \\ -1{,}14456645 & 0{,}046277729 \\ -0{,}43804395 & 0{,}01771123 \\ -1{,}22384085 & -0{,}16282447 \end{bmatrix} * \begin{bmatrix} 0{,}677782 & 0{,}735263 \\ -0{,}735096 & 0{,}677963 \end{bmatrix}$$

$$Row\ Zero\ Mean\ Data = \begin{bmatrix} 0{,}68982091 & 0{,}49010659 \\ -1{,}30965445 & -1{,}21027104 \\ 0{,}38988461 & 0{,}99023530 \\ 0{,}08997215 & 0{,}29006940 \\ 1{,}28966174 & 1{,}0902419 \\ 0{,}48986404 & 0{,}79018421 \\ 0{,}18995952 & -0{,}31007988 \\ -0{,}80978511 & -0{,}81018277 \\ -0{,}30991775 & -0{,}31006995 \\ -0{,}70980568 & -1{,}01023386 \end{bmatrix}$$

$$Data\ Baru = \ Row\ Zero\ Mean\ Data + Mean\ Global\ Data$$

$$Data\ Baru = \begin{bmatrix} 0{,}68982091 & 0{,}49010659 \\ -1{,}30965445 & -1{,}21027104 \\ 0{,}38988461 & 0{,}99023530 \\ 0{,}08997215 & 0{,}29006940 \\ 1{,}28966174 & 1{,}0902419 \\ 0{,}48986404 & 0{,}79018421 \\ 0{,}18995952 & -0{,}31007988 \\ -0{,}80978511 & -0{,}81018277 \\ -0{,}30991775 & -0{,}31006995 \\ -0{,}70980568 & -1{,}01023386 \end{bmatrix} + [1{,}81 \quad 1{,}91]$$

$$Data\ Baru = \begin{bmatrix} 2,499821 & 2,400107 \\ 0,500346 & 0,699729 \\ 2,199885 & 2,900235 \\ 1,899972 & 2,200069 \\ 3,099662 & 3,000242 \\ 2,299864 & 2,700184 \\ 1,99996 & 1,59992 \\ 1,000215 & 1,099817 \\ 1,500082 & 1,59993 \\ 1,100194 & 0,899766 \end{bmatrix}$$

8. Mengurangi dimensi N menjadi dimensi K.

   $\sum \lambda_i = \lambda_1 + \lambda_2 = 1,284094 + 0,049239 = 1,333333$

   Jika minimal 90% informasi yang tetap dipertahankan, maka:

   $$\frac{90}{100} * 1,333333 = 1,2$$

   Maka nilai eigen yang terpilih sebagai komponen utama adalah nilai eigen yang memiliki $\lambda \geq 1,2$.

   Maka yang terpilih adalah $\lambda_2 = 1,284094$ sehingga eigen vector nya adalah:

   $$U = \begin{bmatrix} 0,677782 \\ 0,735263 \end{bmatrix}$$

9. Menentukan fitur yang baru:

   $$(k\ eigen\ vector^T * Zero\ Mean^T)^T$$

   $$\left(\begin{bmatrix} 0,677782 \\ 0,735263 \end{bmatrix}_{2x1}^T * \begin{bmatrix} 0,69 & 0,49 \\ -1,31 & -1,21 \\ 0,39 & 0,99 \\ 0,09 & 0,29 \\ 1,29 & 1,09 \\ 0,49 & 0,79 \\ 0,19 & -0,31 \\ -0,81 & -0,81 \\ -0,31 & -0,31 \\ -0,71 & -1,01 \end{bmatrix}_{10x2}^T \right)^T$$

   $$= \left[[0,677782 \quad 0,735263]_{1x2} * \begin{bmatrix} 2,5 & 0,5 & 2,2 & 1,9 & 3,1 & 2,3 & 2,0 & 1,0 & 1,5 & 1,1 \\ 2,4 & 0,7 & 2,9 & 2,2 & 3,0 & 2,7 & 1,6 & 1,1 & 1,6 & 0,9 \end{bmatrix}_{2x10} \right]^T$$

   Hasil perkaliannya terdiri atas dua baris baris pertama dan baris kedua. Kedua baris ini akan digabungkan membentuk menjadi susunan matriks 1x10

   $$= [(0,677782 * 2,5 + 0,735263 * 2,4) \quad (0,677782 * 0,5 + 0,735263 * 0,7) +$$

   $$(0,677782 * 2,2 + 0,735263 * 2,9) \quad (0,677782 * 1,9 + 0,735263 * 2,2)$$

   $$(0,677782 * 3,1 + 0,735263 * 3,0) \quad (0,677782 * 2,3 + 0,735263 * 2,7)$$

$$(0{,}677782 * 2{,}0 + 0{,}735263 * 1{,}6) \quad (0{,}677782 * 1{,}0 + 0{,}735263 * 1{,}1)$$

$$(0{,}677782 * 1{,}5 + 0{,}735263 * 1{,}6) \quad (0{,}677782 * 1{,}1 + 0{,}735263 * 0{,}9)]^T$$

$$= [3{,}4590862 \quad 0{,}8535751 \quad 3{,}6233831 \quad 2{,}9053644 \quad 4{,}3069132$$

$$3{,}5441087 \quad 2{,}5319848 \quad 1{,}4865713 \quad 2{,}1930938 \quad 1{,}4072969]^T$$

$$= \begin{bmatrix} 3{,}4590862 \\ 0{,}8535751 \\ 3{,}6233831 \\ 2{,}9053644 \\ 4{,}3069132 \\ 3{,}5441087 \\ 2{,}5319848 \\ 1{,}4865713 \\ 2{,}1930938 \\ 1{,}4072969 \end{bmatrix}$$

10. Menentukan komponen utama (principal component) hasil ekstraksi.
11. Membangun kembali data yang baru hasil reduksi.