

**Univerzita Pardubice
Fakulta chemicko-technologická**

Katedra analytické chemie

**Metody s latentními proměnnými a
klasifikační metody**

**Ing. Roman Slavík
ŽDB a.s.**

V Bohumíně 14.4.2001

Příklad č.1

Zadání :

Vypočtete algoritmem NIPALS 1. latentní proměnnou z matice A [řádek, sloupec]: A [1,1] = 1, A [2,1] = 2, A [3,1] = 3, A [1,2] = 1, A [2,2] = 2, A [3,2] = 0, A [1,3] = 6, A [2,3] = 4, A [3,3] = 2. Matici před zpracováním standardizujte.

Řešení :

$$\text{zdrojová matice} \quad A = \begin{vmatrix} 1 & 1 & 6 \\ 2 & 2 & 4 \\ 3 & 0 & 2 \end{vmatrix}$$

$$\text{Nejdříve vypočteme vektor aritmetických průměrů} \quad \mathbf{x}^T = \begin{vmatrix} 2 & 1 & 4 \end{vmatrix}$$

$$\text{a vektor směrodatných odchylek} \quad \mathbf{s}^T = \begin{vmatrix} 1 & 1 & 2 \end{vmatrix}$$

$$\text{Standardizací obdržíme matici} \quad A = \begin{vmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & -1 \end{vmatrix}$$

Jako první odhad vektoru $\mathbf{t}_1$ byl zvolen první sloupec standardizované matice :

$$\mathbf{t}_1^T = \begin{vmatrix} -1 & 0 & 1 \end{vmatrix}$$

$$\text{Po dosazení do vztahu} \quad \mathbf{p}_1^T = (\mathbf{t}_1^T \mathbf{t}_1)^{-1} \mathbf{t}_1^T A$$

$$\text{a znormováním podle} \quad \mathbf{p}_1^N = (\mathbf{p}_1^T \mathbf{p}_1)^{-1/2} \mathbf{p}_1$$

získáme počáteční odhad vektoru míry příspěvků odhadu vektoru hlavní komponenty $\mathbf{t}_1$. Dosazením za $\mathbf{t}_1^T = (\mathbf{p}_1^T \mathbf{p}_1)^{-1} \mathbf{p}_1^T A^T$

získáme odhad hlavní komponenty $\mathbf{t}_1$. Iteračním opakováním tohoto postupu dostaneme stabilní rozklad vektorů $\mathbf{t}_1$ a $\mathbf{p}_1$.

Konvergenční kritérium má tvar :
$$\mathbf{d} = (\mathbf{t}_{\text{nové}} - \mathbf{t}_{\text{staré}})^T (\mathbf{t}_{\text{nové}} - \mathbf{t}_{\text{staré}}) (\mathbf{t}_{\text{nové}}^T - \mathbf{t}_{\text{nové}}^T)^1$$

 Výpočet se ukončí když
$$\mathbf{d} / N < 10^{-10}$$

Výpočet vektoru p_1 :

$$\begin{aligned}
 t_1^T t_1 &= \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \end{bmatrix} \xrightarrow{\text{inverze}} \begin{bmatrix} 0,5 \end{bmatrix} \xrightarrow{*t_1^T} \begin{bmatrix} 0,5 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -0,5 & 0 & 0,5 \end{bmatrix} \xrightarrow{*X} \\
 &\xrightarrow{*X} \begin{bmatrix} -0,5 & 0 & 0,5 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & -1 & -1 \end{bmatrix} \\
 p_1^T &= \begin{bmatrix} 1 & -0,5 & -1 \end{bmatrix}
 \end{aligned}$$

Normování vektoru p_1 :

$$\begin{aligned}
 (p_1^T p_1)^{-1/2} &= \left(\begin{bmatrix} 1 & -0,5 & -1 \end{bmatrix} * \begin{bmatrix} 1 \\ -0,5 \\ -1 \end{bmatrix} \right)^{-1/2} = (2,25)^{-1/2} = 0,6667 \xrightarrow{*p_1} 0,6667 * \begin{bmatrix} 1 & -0,5 & -1 \end{bmatrix} \\
 p_1^N &= \begin{bmatrix} 0,6667 & -0,3333 & -0,6667 \end{bmatrix}
 \end{aligned}$$

Výpočet t_1 :

$$\begin{aligned}
 p_1^T p_1 &= \begin{bmatrix} 0,6667 & -0,3333 & -0,6667 \end{bmatrix} * \begin{bmatrix} 0,6667 \\ -0,3333 \\ -0,6667 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \xrightarrow{\text{inverze}} \begin{bmatrix} 1 \end{bmatrix} \xrightarrow{*p_1^T} \\
 &\xrightarrow{*p_1^T} \begin{bmatrix} 0,6667 & -0,3333 & -0,6667 \end{bmatrix} \xrightarrow{*X} \begin{bmatrix} 0,6667 & -0,3333 & -0,6667 \end{bmatrix} * \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix} \\
 t_1^T &= \begin{bmatrix} -1,3333 & 0,3333 & 1,6667 \end{bmatrix}
 \end{aligned}$$

Výpočet se opakuje. Jednotlivé vypočítané vektory jsou uvedeny v tabulce č.1.

Metody s latentními proměnnými a klasifikační metody

Tabulka č. 1 - Vypočítané vektory latentních proměnných a zátěží.

Číslo opakování	Vektor zátěží (p1)	Normovaný vektor zátěží (p1N)	Vektor latentních hodnot
1	1 -0,5 -1	0,6667 -0,3333 -0,6667	-1,3333 -0,3334 1,6667
2	0,6750 -0,4500 -0,6750	0,6396 -0,4264 -0,6396	-1,2792 -0,4264 1,7056
3	0,6314 -0,4510 -0,6314	0,6312 -0,4508 -0,6312	-1,2623 -0,4508 1,7132
4	0,6289 -0,4574 -0,6289	0,6288 -0,4573 -0,6288	-1,2577 -0,4573 1,7150
5	0,6282 -0,4591 -0,6282	0,6282 -0,4591 -0,6282	-1,2564 -0,4591 1,7154
6	0,6280 -0,4595 -0,6280	0,6280 -0,4595 -0,6280	-1,2561 -0,4596 1,7157
7	0,6279 -0,4596 -0,6279	0,6280 -0,4597 -0,6280	-1,2559 -0,4597 1,7156
8	0,6280 -0,4597 -0,6280	0,6280 -0,4597 -0,6280	-1,2559 -0,4597 1,7156

Po 8 krocích obdržíme :

$$\text{zátěžový vektor} \quad p_1^T = \begin{vmatrix} 0,6279675 & -0,4596886 & -0,6279675 \end{vmatrix}$$

$$\text{vektor latentních proměnných} \quad t_1^T = \begin{vmatrix} -1,255935 & -0,4596886 & 1,7156236 \end{vmatrix}$$

Pronásobením a odečtením dle

$$E_1 = A - t_1 p_1^T$$

získáme matici nevysvětlené variability

$$E_1 = \begin{vmatrix} -0,21131 & -0,57734 & 0,211314 \\ 0,28867 & 0,788686 & -0,28867 \\ -0,07736 & -0,21135 & 0,077356 \end{vmatrix}$$

Závěr:

Algoritmem NIPALS byla spočtena první latentní proměnná.

Příklad č.2

Zadání :

S použitím vhodných kritérií určete nezbytný počet latentních proměnných, bylo-li z dat určeno: PRESS(0) = S(0) = 100, PRESS(1) = 20, S(1) = 10, PRESS(2) = 3,5; S(2) = 3,4; PRESS(3) = 3,45; S(3) = 3,39.

Řešení :

PRESS(0)=S(0) - suma x_{ki}^2 ze zdrojové matice

PRESS(P) – suma e_{ki}^2 z matice E^{pred}

S(P) – suma e_{ki}^2 z matice E

K určení nezbytného počtu latentních proměnných použijeme Woldovo kritérium založené na poměru hodnoty PRESS(P) a hodnoty S(P-1).

$$\text{PRESS(P)/S(P-1)}$$

$$P = 1 : \text{PRESS(1)/S(0)} = 0,2$$

$$P = 2 : \text{PRESS(2)/S(1)} = 0,35$$

$$P = 3 : \text{PRESS(3)/S(2)} = 1,01 \quad 1,01 > 0,95$$

Závěr:

Woldovým kritériem byl stanoven počet latentních proměnných na základě výše uvedených hodnot na 3. Čtvrtá latentní proměnná není třeba, protože hodnota kritéria vyšla větší (1,01) než 0,95.

Příklad č.3**Zadání :**

Odhadněte hodnotu chybějícího prvku $A[2,2]$, jestliže výpočtem z nekompletní matice byly určeny vektory p : 0,541 0,423 0,514 0,514 t : -1,340 -0,735 2,076

Řešení :

Prvky zdrojové matice odpovídající p -té latentní proměnné lze rekonstruovat podle vzorce

$$A_p^{\text{pred}} = t_p p_p^T$$

vektor p		0,541	0,423	0,514	0,514	
vektor t		-1,340	-0,735	2,076		
matice A^{pred}		-0,7249	-0,5668	-0,6888	-0,6888	
		0,3976	-0,3109	-0,3778	-0,3778	
		1,1231	0,8782	1,0671	1,0671	

$$A[2,2] = -0,311$$

Závěr:

První odhad chybějící hodnoty činí -0,311.

Příklad č.4**Zadání :**

Výpočtem metodou PCA byly určeny vektory p_1 : 0,012 0,458 -0,352 0,987 p_2 : 0,926 -0,238 0,872 -0,115. Vypočítejte komunalitu a vyberte sloupec, který nejlépe charakterizuje celou matici.

Řešení :

Prvek matice zátěží p_{ip} příslušejícímu i -tému sloupci zdrojové matice je mírou variability tohoto sloupce popsané p -tou latentní proměnnou. Podíl variability daného sloupce popsané společnými latentními proměnnými lze pak vyjádřit jako součet příspěvků jednotlivých latentních proměnných, tedy

$$h_i^2 = p^N (p^N)^T = \sum_{p=1}^P (p_{ip}^N)^2$$

kde h_i^2 je komunalita pro i -tý sloupec
 P ... počet latentních proměnných
 M ... počet sloupců zdrojové matice

$$p_{ip}^N \text{ jsou zátěže normované podle vztahu: } p_{ip}^N = \frac{p_{ip}}{\sum_{i=1}^M p_{ip}^2}$$

Výpočet normovaných zátěží

vektor p_1		0,012	0,458	-0,352	0,987	
vektor p_2		0,926	-0,238	0,872	-0,115	

Vypočítáme normované zátěže pro jednotlivé sloupce při dvou latentních proměnných.

$$p_{11}^N = \frac{p_{11}}{\sum_{i=1}^M p_{i1}^2} = \frac{0,012}{0,012^2 + 0,458^2 + (-0,352)^2 + 0,987^2} = \frac{0,012}{1,3080} = 0,0092$$

$$p_{21}^N = \frac{p_{21}}{\sum_{i=1}^M p_{i1}^2} = \frac{0,458}{0,012^2 + 0,458^2 + (-0,352)^2 + 0,987^2} = \frac{0,458}{1,3080} = 0,3502$$

$$p_{31}^N = \frac{p_{31}}{\sum_{i=1}^M p_{i1}^2} = \frac{-0,352}{0,012^2 + 0,458^2 + (-0,352)^2 + 0,987^2} = \frac{-0,352}{1,3080} = -0,2691$$

$$p_{41}^N = \frac{p_{41}}{\sum_{i=1}^M p_{i1}^2} = \frac{0,987}{0,012^2 + 0,458^2 + (-0,352)^2 + 0,987^2} = \frac{0,987}{1,3080} = 0,7546$$

$$p_{12}^N = \frac{p_{12}}{\sum_{i=1}^M p_{i2}^2} = \frac{0,926}{0,926^2 + (-0,238)^2 + 0,872^2 + (-0,115)^2} = \frac{0,926}{1,6877} = 0,5487$$

$$p_{22}^N = \frac{p_{22}}{\sum_{i=1}^M p_{i2}^2} = \frac{-0,238}{0,926^2 + (-0,238)^2 + 0,872^2 + (-0,115)^2} = \frac{-0,238}{1,6877} = -0,1410$$

$$p_{32}^N = \frac{p_{32}}{\sum_{i=1}^M p_{i2}^2} = \frac{0,872}{0,926^2 + (-0,238)^2 + 0,872^2 + (-0,115)^2} = \frac{0,872}{1,6877} = 0,5167$$

$$p_{42}^N = \frac{p_{42}}{\sum_{i=1}^M p_{i2}^2} = \frac{-0,115}{0,926^2 + (-0,238)^2 + 0,872^2 + (-0,115)^2} = \frac{-0,115}{1,6877} = -0,0681$$

Výpočet komunalit

$$h_1^2 = \sum_{p=1}^P (p_{1p}^N)^2 = p_{11}^2 + p_{12}^2 = 0,0092^2 + 0,54487^2 = 0,2970$$

$$h_2^2 = \sum_{p=1}^P (p_{2p}^N)^2 = p_{21}^2 + p_{22}^2 = 0,3501^2 + (-0,1410)^2 = 0,1430$$

$$h_4^2 = \sum_{p=1}^P (p_{4p}^N)^2 = p_{41}^2 + p_{42}^2 = 0,7546^2 + (-0,0681)^2 = 0,5741$$

Závěr:

Platí pravidlo, že čím je komunalita příslušného sloupce větší, tím má sloupec vlastnosti společné s ostatními sloupci zdrojové matice. V našem případě čtvrtý sloupec nejlépe charakterizuje zdrojovou matici.

Příklad č.5

Zadání :

Vysvětlete, proč vysvětlená variabilita je při výpočtu metodou FA vždy nižší, než při výpočtu metodou PCA.

Řešení :

Rozdíl spočívá ve způsobu rozkladu variability zdrojové matice.

V případě **metody PCA** je model postaven na $S^2 = H^2$, (S^2 - diagonální matice rozptylů manifestních proměnných, H^2 - diagonální matice komunalit), který předpokládá, že lze variabilitu zdrojové matice reprodukovat přesně .

U **metody FA** platí, že variabilita je rozložena do dvou složek $S^2 = H^2 + L^2$, (L^2 - matice jedinečností, která představuje část variability nevysvětlitelné společnými faktory). Tím, že u metody FA předpokládáme existenci nevysvětlené variability dospíváme vždy k nižším hodnotám vysvětlené variability než u metody PCA.

Příklad č.6

Zadání :

Výpočtem metodou kanonických korelací bylo zjištěno:

$$0,297X_1 + 0,298X_2 + 0,050X_3 + 0,256X_4 = 0,493Y_1 - 0,213Y_2$$

$$r_1 = 0,830$$

$$0,006X_1 - 0,115X_2 + 0,950X_3 + 0,056X_4 = 0,493Y_1 + 0,213Y_2$$

$$r_1 = 0,512$$

Vypočítejte skupinový korelační koeficient a interpretujte výsledky.

Řešení :

Skupinový korelační koeficient R vypočteme ze vztahu

$$R^2 = 1 - (1 - r_1^2)(1 - r_2^2)$$

$$R^2 = 0,770$$

Závěr:

Matice X je tvořena 4 sloupci a matice Y dvěma. Skupinový korelační koeficient má hodnotu 0,770, což znamená, že 77 % variability dat jsme vysvětlili kanonickými korelačními koeficienty.

S růstem proměnných X_1 , X_2 a X_4 , které mají stejný vliv, klesá význam proměnné Y_2 a roste významně proměnná Y_1 . S nadbytkem proměnné X_3 se zvyšuje proměnná Y_2 .

Příklad č.7

Zadání :

Při monitoringu vstupních vod do naší akciové společnosti jsou v chemické laboratoři analyzovány vzorky těchto odebraných vod. U těchto povrchových vod se v naší laboratoři kromě jiných stanovují tyto parametry : Ph, Vodivost, NL, RL, Tvrdost, $-NH_4$, $-NO_3$, $-NO_2$, Cl, CHSK Cr, Ba, Fe, Sr, BSK5. Parametr AOX je stanovován externí laboratoří v Paskově. Na datech v tabulce se pokuste najít vztah mezi AOX a parametry stanovovanými v naší laboratoři. Kvalita povrchové vody by byla známa již při stanovení parametrů v naší laboratoři.

Metody s latentními proměnnými a klasifikační metody

VZ	PH	VODIV	NL	RL	TVRD	-NH4	-NO3	-NO2	CL	CHSK	CR	BA	FE	SR	AOX	BSK5
10	7,470	296,000	10,000	1716,000	23,100	0,370	23,400	0,610	799,000	21,000	0,286	0,534	1,883	0,496	7,200	
11	7,580	67,000	50,000	380,500	13,900	5,600	15,100	0,140	45,200	17,800	0,077	0,659	0,298	0,038	4,000	
12	7,660	45,100	17,000	190,000	9,500	0,420	21,000	0,270	31,300	16,400	0,075	0,991	0,251	0,026	6,300	
20	7,950	206,000	32,000	1473,000	21,300	7,600	11,100	0,500	418,000	21,000	0,277	0,635	1,318	0,233	4,200	
26	7,870	238,000	69,000	1413,000	23,500	9,950	11,700	0,520	521,000	21,800	0,343	1,500	1,613	0,458	3,500	
356	7,280	161,300	37,000	815,500	18,100	3,850	17,000	0,060	310,000	20,000	0,166	0,529	0,933	0,208	3,200	
359	6,980	283,000	24,000	1615,000	21,600	0,660	22,700	0,420	757,000	23,200	0,201	0,396	1,640	0,315	0,500	
360	6,820	49,000	33,000	336,500	9,900	0,950	30,300	0,270	34,000	28,100	0,059	0,841	0,180	0,036	6,300	
361	7,110	55,300	12,000	385,000	11,900	0,310	35,600	0,190	44,700	17,190	0,067	0,498	0,168	0,040	2,400	
362	7,330	160,800	43,000	995,500	22,100	2,090	17,600	0,300	310,000	28,590	0,159	1,690	0,865	0,241	5,100	
363	7,160	44,100	49,000	323,000	9,500	0,480	31,500	0,280	31,000	29,580	0,050	0,744	0,196	0,022	1,500	
742	7,690	171,100	24,000	901,000	19,000	5,650	10,800	0,510	344,000	22,000	0,224	0,323	1,096	0,261	3,500	
745	7,340	209,000	25,000	1218,000	17,100	0,750	15,900	0,280	516,000	18,000	0,196	0,244	1,406	0,454	2,500	
746	7,280	62,000	5,000	170,500	13,700	0,480	19,900	0,160	34,400	16,000	0,078	0,430	0,296	0,047	3,500	
747	7,460	56,300	8,000	369,000	12,000	0,400	36,100	0,170	41,300	16,000	0,098	0,148	0,219	0,034	2,500	
748	7,660	174,600	23,000	950,000	19,600	5,700	10,000	0,500	310,000	12,000	0,213	0,315	1,202	0,121	3,500	
749	7,460	33,600	4,000	318,000	7,000	0,650	22,200	0,120	20,700	16,000	0,065	0,236	0,188	0,020	2,000	
1155	7,770	164,700	14,000	862,000	18,900	0,680	8,700	0,250	313,000	38,650	0,207	0,682	1,094	0,156	6,200	
1158	6,890	99,000	33,000	526,500	9,800	0,120	13,800	0,060	215,000	20,310	0,260	0,730	0,922	0,133	6,600	
1159	6,790	60,700	10,000	686,000	13,200	0,400	21,400	0,280	34,800	24,440	0,066	0,639	0,304	0,035	6,700	
1160	6,780	60,000	5,000	462,500	12,000	0,640	32,300	1,630	45,200	11,950	0,104	7,750	0,233	0,020	3,100	
1161	7,330	165,000	60,000	1530,000	19,000	1,170	9,500	0,370	313,000	30,870	0,258	1,150	1,148	0,132	5,700	
1162	6,850	27,800	13,000	273,000	5,600	0,500	18,600	0,060	13,900	21,030	0,048	0,309	0,154	0,020	5,100	
3663	7,750	198,800	45,000	1491,000	20,700	0,280	8,000	0,290	452,000	32,000	0,261	0,594	1,312	0,092	5,300	
3991	7,950	223,000	39,000	1369,000	25,480	0,100	7,400	0,090	532,000	21,000	0,288	8,940	1,578	0,174	6,000	

Příklad č.8

Zadání :

Jeden objekt je charakterizován metrickými znaky (2,10), druhý (3,8), třetí (4,9), čtvrtý (10,4) a pátý (11,5). Vypočítejte matici vzdáleností v Euklidově metrice a dokumentujte výpočet shlukování některou z používaných metod. Výsledky interpretejte graficky.

Řešení :

Postupy aglomerativní shlukové analýzy lze neuspořádanou množinu objektů charakterizovaných náhodnými vektory postupně uspořádat do tříd až ke konečnému stavu, kdy se všechny objekty spojí do jediné třídy. Výše uvedené vlastnosti objektů (blízkost či podobnost) posuzujeme podle vzdálenosti objektů v p-rozměrném prostoru znaků nejjednodušším typem Euklidovy metriky, která je definovaná vztahem $d_E(X_k, X_i) = [\sum (x_{kp} - x_{ip})^2 \text{ pro } p = 1..p]^{1/2}$.

Vzdálenost určená podle zásady nejbližšího souseda (nejbližší jsou ty třídy, které mají nejmenší vzdálenost mezi dvěma nejbližšími objekty (objekt můžeme chápat jako třídu)) se vypočte podle vzorce $D_{NN} = (S_r, S_s) = \min. d(X_k, X_l)$.

Matice vzdáleností

X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅
X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅
X ₄₁	X ₄₂	X ₄₃	X ₄₄	X ₄₅
X ₅₁	X ₅₂	X ₅₃	X ₅₄	X ₅₅
0				

Metody s latentními proměnnými a klasifikační metody

Matice vzdáleností	2,236	0				
	2,236	1,414	0			
	10	8,062	7,810	0		
	10,296	8,544	8,062	1,414	0	

Hierarchical Cluster Analysis

Complete Linkage Euclidean Distance
Amalgamation Steps

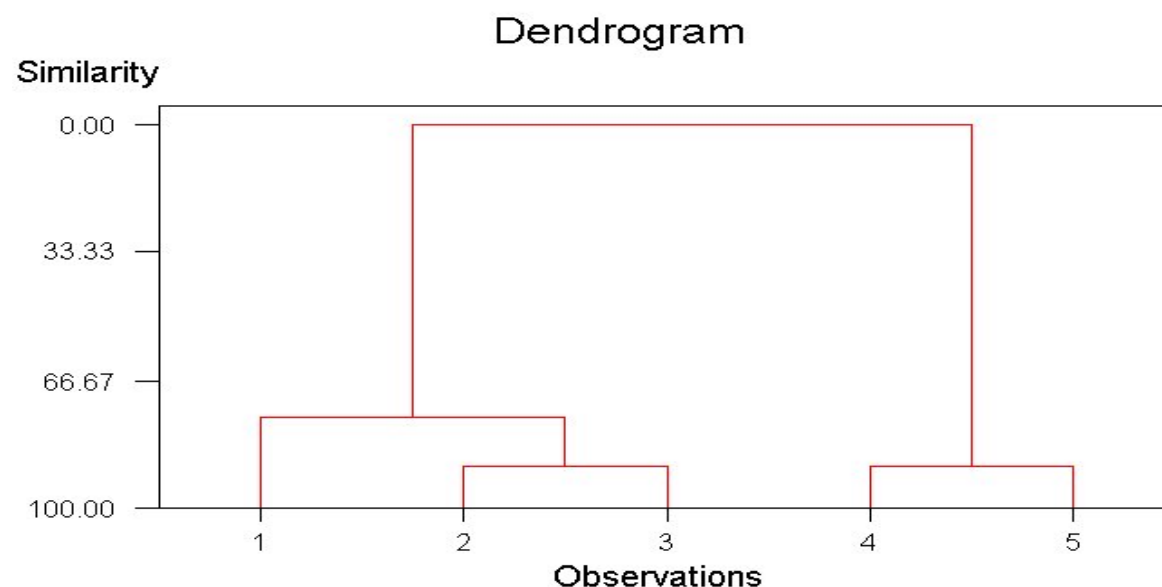
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	4	89.07	2.072	2 3	2	2
2	3	88.96	2.093	4 5	4	2
3	2	76.14	4.523	1 2	1	3
4	1	0.00	18.960	1 4	1	5

Final Partition

Number of clusters: 1

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	5	377.749	8.512	10.697

Grafické znázornění jednotlivých objektů



Z grafického pohledu vyplývá, že objekty postupem nejbližšího souseda vytvářejí nejdříve dva shluky a to 2-3 a 4-5 dalším postupem je objekt č.1 připojen ke shluku 2-3 a následně tento shluk vytvoří jeden shluk se shlukem 4-5.

Závěr:

Byly stanoveny vzdálenosti mezi jednotlivými objekty formou matice vzdáleností a metodou shlukové analýzy klasifikovány objekty postupem nejbližšího souseda.

Příklad č.9

Monitorování kvality podzemních i povrchových vod.

Metody s latentními proměnnými a klasifikační metody

Zadání :

Monitoring je prováděn na území zahrnující hutní odval (vlastník ŽDB), skládku tuhých průmyslových odpadů (vlastník ŽDB) a městskou skládku komunálního odpadu (vlastník BMS), včetně přilehlých území. Veškeré terénní, vzorkovací a laboratorní práce jsou finančně i časově velmi náročné. Rozhodněte, zda je možné redukovat počet odběrových míst a charakterizovat jednotlivé vrty menším počtem stanovovaných ukazatelů.

Data :

Bylo vzorkováno 20 vrtů a jedna voda povrchová.

VZOREK	OZN	PH	VODIVOST	F	CL	CHSK CR	B	BA	CU	FE	MN	ZN	SN
3438	OD 1	6,980	250,000	1,000	386,100	62,000	0,131	0,065	0,009	27,300	17,600	0,018	0,650
3439	OD 3	7,050	364,000	2,550	772,200	49,000	3,470	0,059	0,008	6,160	1,710	0,009	0,394
3440	OD 4	7,020	249,000	1,390	386,100	53,000	4,200	0,240	0,038	62,000	1,570	0,046	0,071
3441	OD 5	8,110	366,000	3,600	702,000	30,000	4,100	0,069	0,014	0,617	0,087	0,007	0,400
3442	OD 7	6,570	344,000	1,150	737,100	52,000	0,198	0,484	0,013	93,200	5,970	0,010	0,070
3443	OD 8	7,760	351,000	2,200	666,900	68,000	2,400	0,103	0,004	3,260	3,140	0,022	0,527
3444	PV 32	6,870	215,000	0,480	122,900	32,000	4,270	0,063	0,012	24,300	2,500	0,060	0,438
3445	PV 102	7,700	242,000	5,750	280,800	41,000	6,130	0,068	0,029	2,410	1,580	0,023	0,192
3446	PV 105	7,050	222,000	1,130	80,730	36,000	0,505	0,069	0,165	19,500	3,460	0,377	0,467
3447	PV 106	6,770	271,000	0,480	280,800	170,000	2,740	0,112	0,024	74,700	14,900	0,045	0,682
3448	BC 2	7,600	256,000	0,380	631,800	21,000	0,430	0,340	0,001	1,050	0,363	0,048	0,180
3450	PV 1	6,710	366,000	0,250	912,600	41,000	0,428	0,071	0,054	2,190	4,130	0,045	0,061
3451	PV 2	7,210	196,300	0,900	210,600	35,000	3,860	0,044	0,018	0,177	0,320	0,025	0,049
3452	PV 5	7,520	219,000	0,880	421,200	12,000	0,437	0,068	0,019	0,095	11,850	0,031	0,155
3453	PV 9	6,770	157,700	0,550	175,500	22,000	1,390	0,113	0,073	12,100	1,610	0,112	0,028
3454	PV 10	6,340	348,000	0,290	702,000	23,000	7,340	0,080	0,027	0,643	2,690	0,073	0,097
3455	PV 11	6,290	166,400	0,690	263,200	10,000	0,949	0,121	0,018	12,000	1,900	0,018	0,078
3456	PV 12	5,850	375,000	4,100	1123,000	40,000	0,039	0,607	0,016	96,600	23,650	0,074	0,157
3457	PV 40	5,670	137,100	5,100	421,200	13,000	0,114	0,509	0,024	53,500	4,000	0,028	0,047
3458	PV 103	6,680	282,000	0,510	473,800	43,000	1,375	0,084	0,017	0,984	1,250	0,036	0,084
3459	PV 104	6,560	161,700	0,940	98,280	40,000	0,436	0,131	0,013	34,900	3,750	0,024	0,081

Postup při analýze dat :

- Nejdříve provedeme průzkumovou analýzu dat. Tím odstraníme případné chyby.
- Analýza korelační matice a matice parciálních korelačních koeficientů mezi vlastnostmi.
- Zvolíme metodu. V našem případě zvolíme faktorovou analýzu a metodu hlavních komponent.
- Data standardizujeme, protože nejsou vyjádřena ve stejných jednotkách.
- Vypočteme latentní proměnné
- Určíme počet signifikantních latentních proměnných.
- Analyzujeme matici zátěží a matici latentních proměnných.
- Analyzujeme fyzikálně-chemický smysl latentních proměnných.
- Interpretujeme výsledky vzhledem k cíli analýzy.