

Zastosowanie miar zależności dla wielowymiarowych danych funkcjonalnych

Tomasz Górecki

tomasz.gorecki@amu.edu.pl

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu
Zakład Rachunku Prawdopodobieństwa i Statystyki Matematycznej



- 1 *Wprowadzenie*
- 2 *Wyglądanie danych dyskretnych*
- 3 *Wybór zmiennych w klasyfikacji danych funkcjonalnych*
 - Korelacja odległościowa
 - Wybór zmiennych
 - Przykład
- 4 *Analiza kanoniczna dla wielowymiarowych danych funkcjonalnych oparta na HSIC*
 - Jądrowy współczynnik zgodności wektorów losowych
 - Jądrowy współczynnik zgodności wielowymiarowych procesów losowych
 - Analiza kanoniczna
 - Przykład
- 5 *Literatura*

W ostatnich latach reprezentacja danych za pomocą funkcji lub krzywych zyskuje coraz większe znaczenie. Tego typu dane znane są jako **dane funkcjonalne** (Ramsay & Silverman, 2005). Tego typu dane można spotkać w medycynie, ekonomii, meteorologii i wielu innych dziedzinach nauki. W licznych zastosowaniach praktycznych zachodzi potrzeba użycia metod statystycznych dla obiektów scharakteryzowanych przez wiele cech obserwowanych w wielu punktach czasowych. Takie dane nazywamy **wielowymiarowymi danymi funkcjonalnymi**.

Założmy, że $\mathbf{X} \in L_2^p(I)$ jest procesem losowym, gdzie $L_2(I)$ jest przestrzenią Hilberta funkcji całkowalnych z kwadratem na przedziale I . Dodatkowo założmy, że

$$E(\mathbf{X}(s)) = \mathbf{0}, \quad s \in I.$$

Założmy dalej, że każda składowa X_g procesu \mathbf{X} może być przedstawiona jako suma **skończonej liczby funkcji bazowych** $\{\varphi_e\}$:

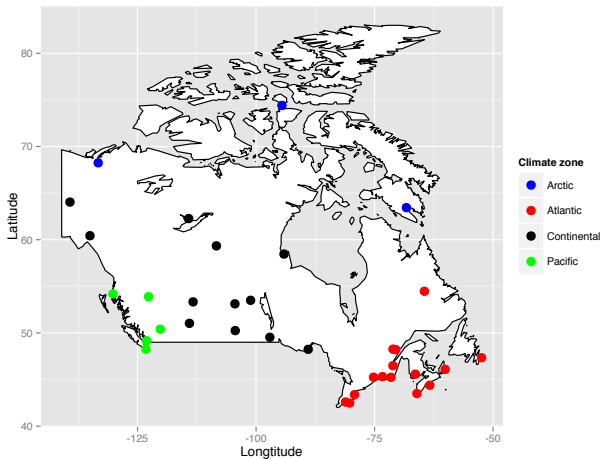
$$X_g(s) = \sum_{e=0}^{E_g} \alpha_{ge} \varphi_e(s), s \in I, g = 1, 2, \dots, p.$$

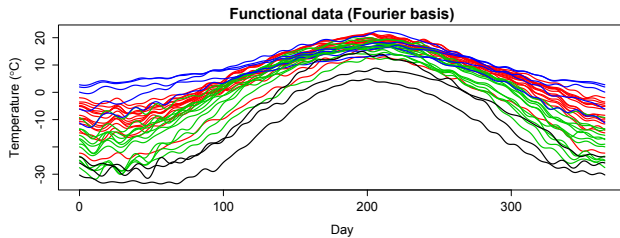
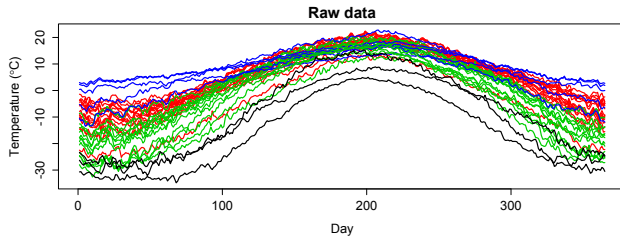
Stopień wygładzenia funkcji X_g zależy od wartości E_g (małe wartości powodują większe wygładzenie).

Wektor α można oszacować na bazie n niezależnych realizacji $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ procesu losowego \mathbf{X} (dane funkcjonalne). Estymator ten oznaczmy przez $\hat{\alpha}$.

Szczegóły metody transformacji danych dyskretnych do danych funkcjonalnych można znaleźć w bibliografii Ramsaya i Silvermana (2005) lub w pracy Góreckiego i innych (2014).

Średnie dzienne temperatury dla 35 stacji meteorologicznych w Kanadzie w latach 1960 – 1994. Dane pochodzą z książki Ramsay'a i Silvermana (2005), *Functional Data Analysis*, Springer, New York.





Funkcja charakterystyczna łącznego rozkładu procesu losowego $\mathbf{X} \in L_2^p(I)$ i wektora losowego $\mathbf{Y} \in \mathbb{R}^q$ ma postać:

$$f_{\mathbf{X}, \mathbf{Y}}(l, \mathbf{m}) = E\{\exp[i \langle l, \mathbf{X} \rangle_p + i \langle \mathbf{m}, \mathbf{Y} \rangle_q]\}.$$

Możemy założyć, że (Ramsay & Silverman (2005)) wektorowa funkcja wagowa l oraz proces \mathbf{X} leżą w tej samej przestrzeni, tzn. funkcja l może być zapisana jako

$$l(t) = \Phi(t)\lambda,$$

gdzie $\lambda \in \mathbb{R}^{K+p}$. Stąd

$$\langle l, \mathbf{X} \rangle_p = \lambda' \alpha,$$

i

$$f_{\mathbf{X}, \mathbf{Y}}(l, \mathbf{m}) = f_{\alpha, \mathbf{Y}}(\lambda, \mathbf{m}),$$

gdzie $f_{\alpha, \mathbf{Y}}(\lambda, \mathbf{m})$ jest funkcją charakterystyczną łącznego rozkładu pary wektorów losowych (α, \mathbf{Y}) .

Bazując na pojęciu **kowariancji odległościowej** pomiędzy dwoma wektorami losowymi (Székely i inni (2007)), możemy wprowadzić **funkcjonalną kowariancję odległościową** pomiędzy procesem losowym \mathbf{X} i wektorem losowym \mathbf{Y} jako liczbę nieujemną $\nu_{\mathbf{X}, \mathbf{Y}}$ zdefiniowaną jako

$$\nu_{\mathbf{X}, \mathbf{Y}} = \sqrt{\frac{1}{C_{K+p} C_q} \int_{\mathbb{R}^{K+p+q}} \frac{|f_{\alpha, \mathbf{Y}}(\lambda, \mathbf{m}) - f_{\alpha}(\lambda) f_{\mathbf{Y}}(\mathbf{m})|^2}{\|\lambda\|_{K+p}^{K+p+1} \|\mathbf{m}\|_q^{q+1}} d\lambda d\mathbf{m}},$$

gdzie

$$C_r = \frac{\pi^{\frac{1}{2}(r+1)}}{\Gamma(\frac{1}{2}(r+1))}$$

oraz $f_{\alpha}(\lambda)$, $f_{\mathbf{Y}}(\mathbf{m})$ są funkcjami charakterystycznymi rozkładów brzegowych.

Funkcjonalna korelacja odległościowa pomiędzy procesem losowym \mathbf{X} i wektorem losowym \mathbf{Y} jest liczbą nieujemną zdefiniowaną jako

$$\mathcal{R}_{\mathbf{X},\mathbf{Y}} = \mathcal{R}_{\alpha,\mathbf{Y}} = \frac{\nu_{\alpha,\mathbf{Y}}}{\sqrt{\nu_{\alpha,\alpha}\nu_{\mathbf{Y},\mathbf{Y}}}}$$

jeżeli $\nu_{\alpha,\alpha}$ i $\nu_{\mathbf{Y},\mathbf{Y}}$ są dodatnie i zero w przeciwnym przypadku.
Dla rozkładów ze skończonymi pierwszymi momentami mamy

$$0 \leq \mathcal{R}_{\mathbf{X},\mathbf{Y}} \leq 1$$

i $\mathcal{R}_{\mathbf{X},\mathbf{Y}} = 0$ wtedy i tylko wtedy gdy \mathbf{X} i \mathbf{Y} są niezależne.

Musimy teraz oszacować $\mathcal{R}_{\mathbf{X}, \mathbf{Y}}$ na podstawie $(\hat{\alpha}_k, y_k)$, $k = 1, 2, \dots, n$.

Niech

$$\bar{\alpha} = \frac{1}{n} \sum_{i=1}^n \hat{\alpha}_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$\tilde{\alpha}_k = \hat{\alpha}_k - \bar{\alpha}, \quad \tilde{y}_k = y_k - \bar{y}, \quad k = 1, 2, \dots, n$$

oraz

$$\mathbf{A} = (a_{kl}), \quad \mathbf{B} = (b_{kl}), \\ \tilde{\mathbf{A}} = (A_{kl}), \quad \tilde{\mathbf{B}} = (B_{kl}),$$

gdzie

$$a_{kl} = \|\hat{\alpha}_k - \hat{\alpha}_l\|_{K+p}, \quad b_{kl} = \|\mathbf{y}_k - \mathbf{y}_l\|_q, \\ A_{kl} = \|\tilde{\alpha}_k - \tilde{\alpha}_l\|_{K+p}, \quad B_{kl} = \|\tilde{y}_k - \tilde{y}_l\|_q, \quad k, l = 1, 2, \dots, n.$$

Zachodzą związki

$$\tilde{\mathbf{A}} = \mathbf{H}\mathbf{A}\mathbf{H} \text{ oraz } \tilde{\mathbf{B}} = \mathbf{H}\mathbf{B}\mathbf{H},$$

gdzie

$$\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$$

jest macierzą centrującą.

Niech $\tilde{\mathbf{A}} \circ \tilde{\mathbf{B}} = (A_{kl}B_{kl})$ będzie **iloczynem Hadamarda** macierzy $\tilde{\mathbf{A}}$ oraz $\tilde{\mathbf{B}}$.

Na podstawie wyników Székely'ego i innych (2007), mamy

$$\hat{\nu}_{\mathbf{X},\mathbf{Y}}^2 = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n A_{kl} B_{kl}.$$

Funkcjonalna korelacja odległościowa z próby jest zatem zdefiniowana jako

$$\hat{\mathcal{R}}_{\mathbf{X},\mathbf{Y}} = \hat{\mathcal{R}}_{\hat{\alpha},\mathbf{Y}} = \frac{\hat{\nu}_{\hat{\alpha},\mathbf{Y}}}{\sqrt{\hat{\nu}_{\hat{\alpha},\hat{\alpha}} \hat{\nu}_{\mathbf{Y},\mathbf{Y}}}}$$

jeżeli $\hat{\nu}_{\hat{\alpha},\hat{\alpha}}$ oraz $\hat{\nu}_{\mathbf{Y},\mathbf{Y}}$ są dodatnie i zero w przeciwnym przypadku.

Założmy, że jedynie mała liczba zmiennych objaśniających ma wpływ na \mathbf{Y} . Wybieramy zbiór istotnych zmiennych objaśniających, dla których miara $\hat{\mathcal{R}}_{\mathbf{X},\mathbf{Y}}$ jest duża.

Wykorzystaliśmy funkcjonalną korelację odległościową ponieważ umożliwia ona zbadanie zależności pomiędzy \mathbf{Y} i \mathbf{X} bez względu na to czy jest ona liniowa czy nieliniowa.

Twierdzenie (Kong i inni (2015))

Niech $\mathbf{X}, \mathbf{Z} \in \mathbb{R}^p$ i $\mathbf{Y} \in \mathbb{R}^q$, oraz załóżmy że wektor losowy \mathbf{Z} jest niezależny od wektorów losowych (\mathbf{X}, \mathbf{Y}) . Wtedy

$$\nu_{(\mathbf{X}, \mathbf{Z}), \mathbf{Y}}^2 \leq \nu_{\mathbf{X}, \mathbf{Y}}^2.$$

Zatem jeżeli n jest dostatecznie duże mamy

$$\hat{\nu}_{(\mathbf{X}, \mathbf{Z}), \mathbf{Y}}^2 \leq \hat{\nu}_{\mathbf{X}, \mathbf{Y}}^2,$$

przy założeniu niezależności pomiędzy (\mathbf{X}, \mathbf{Y}) i \mathbf{Z} .

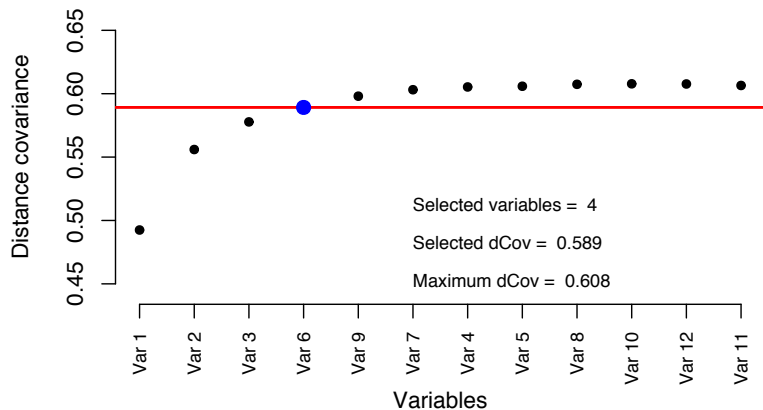
Zastosowaliśmy powyższe twierdzenie jako **regułę stopu** w procedurze wyboru zmiennych. Procedura składa się z następujących kroków:

- 1 Wyznaczamy brzegową kowariancję odległościową dla X_k , $k = 1, \dots, p$ i Y .
- 2 Porządkujemy zmienne w kolejności malejących kowariancji odległościowych. Oznaczmy uporządkowane zmienne objaśniające jako $X_{(1)}, X_{(2)}, \dots, X_{(p)}$. Zaczynamy od $\mathbf{X}_S = \{X_{(1)}\}$.
- 3 Dla k od 2 do p , dodajemy $X_{(k)}$ do \mathbf{X}_S jeżeli $\hat{\nu}_{\mathbf{X}_S, Y}^2$ nie maleje. Zatrzymujemy procedurę w przeciwnym razie.

Jako przykład rzeczywisty użyliśmy zbioru danych **Japanese Vowels**, który jest dostępny w repozytorium: UCI Machine Learning Repository (Lichman (2013)). Dziewięciu mężczyzn wymawiało po kolei dwie japońskie samogłoski /ae/. Dla każdej wymowy, zastosowano 12-stopniową predykcję liniową aby otrzymać dyskretne szeregi czasowe z 12 współczynnikami cepstrum (odwrotna transformata Fouriera). Oznacza to, że każda wymowa to szereg czasowy o długości 7-29 i w każdym punkcie czasowym mamy 12 cech (współczynników). W sumie dysponujemy 640 szeregami, które przynależą do jednej z 9 klas (osoby wymawiające samogłoski) o liczebnościach odpowiednio: 61, 65, 118, 74, 59, 54, 70, 80, 59. Ponieważ szeregi są różnej długości należy je wyrównać (Górecki & Łuczak (2015)).

Podczas procesu wygładzania użyliśmy **bazy Fouriera** o 5 składowych. W kolejnym kroku zastosowaliśmy opisaną wcześniej procedurę wyboru zmiennych (procedura była zatrzymywana jeżeli wzrost miary kowariancji był mniejszy niż 0.01). W taki sposób otrzymaliśmy 4 zmienne.

Przykład – Japońskie samogłoski



Przykład – Japońskie samogłoski

Następnie zastosowaliśmy wybrane metody klasyfikacji do danych zredukowanych i pełnych. Wykorzystaliśmy 10-krokovą metodę sprawdzania krzyżowego (10CV) w celu oszacowania aktualnego poziomu błędu klasyfikacji.

Klasyfikator	DCOV (4)	Wszystkie zmienne (12)
LDC	93.60	99.37
Regresja logistyczna	91.06	97.97
kNN ($k = 1, \dots, 8$)	90.94	96.71
Naiwny Bayes (normal)	90.77	95.50
Naiwny Bayes (kernel)	90.15	94.34
QDC	89.85	Grupy zbyt małe

Zgodnie z oczekiwaniami błąd klasyfikacji wzrósł po redukcji danych. Jednakże, wydaje się, że ten wzrost nie jest zbyt duży. Szczególnie interesujący jest przypadek QDC. Dla tego klasyfikatora nie mamy dostatecznie dużo obserwacji, aby oszacować macierze kowariancji w klasach dla pełnych danych. Jednakże, jeśli ograniczymy się do 4 zmiennych możemy wykonać procedurę.

Niech $\mathbf{X} \in \mathbb{R}^p$ i $\mathbf{Y} \in \mathbb{R}^q$ będą **wektorami losowymi**. Niech

$$k : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$$

będzie rzeczywistą funkcją ciągłą, zwaną **jądrem** i niech $\mathbf{K} = (k_{ij})$, gdzie $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ będzie **macierzą jądrową**, $i, j = 1, \dots, n$. Dalej będziemy przyjmować, że k jest **jądrem gaussowskim**, tj.

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\lambda \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \lambda > 0.$$

Dla każdej macierzy jądrowej $\mathbf{K} \in \mathbb{R}^{n \times n}$, przez **jądrową macierz scentrowaną** rozumiemy nieujemnie określoną macierz postaci

$$\tilde{\mathbf{K}} = \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \mathbf{K} \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right).$$

Niech

$$\langle \mathbf{A}, \mathbf{B} \rangle_F = \text{tr}(\mathbf{A}^T \mathbf{B})$$
$$\|\mathbf{A}\|_F = (\langle \mathbf{A}, \mathbf{A} \rangle_F)^{1/2}$$

będą odpowiednio **iloczynem i normą Frobeniusa** macierzy $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$.

Cortes i inni (2012) zdefiniowali **współczynnik zgodności między macierzami jądrowymi (KTA)** \mathbf{K}_x i \mathbf{K}_y postaci

$$\rho(\mathbf{K}_x, \mathbf{K}_y) = \frac{\langle \tilde{\mathbf{K}}_x, \tilde{\mathbf{K}}_y \rangle_F}{\|\tilde{\mathbf{K}}_x\|_F \cdot \|\tilde{\mathbf{K}}_y\|_F}.$$

Niech \mathbf{X}_s i \mathbf{Y}_t będą procesami losowymi o reprezentacji opisanej powyżej.
Dla procesu \mathbf{X}_s jądro gaussowskie jest równe

$$k_{\mathbf{X}_s}(\mathbf{x}_i(s), \mathbf{x}_j(s)) = k_{\alpha}(\alpha_i, \alpha_j).$$

Podobnie

$$k_{\mathbf{Y}_t}(\mathbf{y}_i(t), \mathbf{y}_j(t)) = k_{\beta}(\beta_i, \beta_j).$$

A to oznacza, że

$$\rho(K_{\mathbf{X}_s}, K_{\mathbf{Y}_t}) = \rho(K_{\alpha}, K_{\beta}) = \frac{\langle \tilde{\mathbf{K}}_{\alpha}, \tilde{\mathbf{K}}_{\beta} \rangle_F}{\|\tilde{\mathbf{K}}_{\alpha}\|_F \|\tilde{\mathbf{K}}_{\beta}\|_F}.$$

Zatem współczynnik zgodności procesów losowych Y_s i Y_t można wyrazić przez współczynnik zgodności wektorów losowych α i β występujących w ich reprezentacji.

W klasycznej analizie korelacji kanonicznych jesteśmy zainteresowani związkiem pomiędzy dwoma wektorami losowymi \mathbf{X} i \mathbf{Y} . W przypadku funkcjonalnym jesteśmy zainteresowani związkiem pomiędzy procesami stochastycznymi \mathbf{X}_s i \mathbf{Y}_t . Funkcjonalne zmienne kanoniczne U i V dla procesów stochastycznych \mathbf{X}_s i \mathbf{Y}_t zdefiniowane są następująco:

$$U = \langle \mathbf{u}(s), \mathbf{X}(s) \rangle = \int_{I_1} \mathbf{u}^T(s) \mathbf{X}(s) ds,$$

$$V = \langle \mathbf{v}(t), \mathbf{Y}(t) \rangle = \int_{I_2} \mathbf{v}^T(t) \mathbf{Y}(t) dt,$$

gdzie funkcje wektorowe $\mathbf{u}(s)$ i $\mathbf{v}(t)$ nazywane są **wektorowymi funkcjami wagowymi** i są postaci

$$\mathbf{u}(s) = \Phi_1(s)\mathbf{u}, \quad \mathbf{v}(t) = \Phi_2(t)\mathbf{v}.$$

Klasycznie funkcje wagowe $\mathbf{u}(s)$ i $\mathbf{v}(t)$ są wybierane tak, aby zmaksymalizować współczynnik korelacji

$$\rho = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U) \text{Var}(V)}},$$

przy warunku $\text{Var}(U) = \text{Var}(V) = 1$.

Współczynnik korelacji pomiędzy zmiennymi U i V jest teraz zastępowany przez współczynnik KTA pomiędzy macierzami jądrowymi $\mathbf{K}_{\mathbf{u}(s)}$ i $\mathbf{K}_{\mathbf{v}(t)}$.

Następnie funkcje wagowe $\mathbf{u}(s)$ i $\mathbf{v}(t)$ są wybierane tak, aby zmaksymalizować współczynnik

$$\hat{\rho}(\mathbf{u}(s), \mathbf{v}(t)) = \frac{\text{tr}(\mathbf{K}_{\mathbf{u}(s)} \tilde{\mathbf{K}}_{\mathbf{v}(t)})}{\sqrt{\text{tr}(\mathbf{K}_{\mathbf{u}(s)} \tilde{\mathbf{K}}_{\mathbf{u}(t)}) \text{tr}(\mathbf{K}_{\mathbf{v}(s)} \tilde{\mathbf{K}}_{\mathbf{v}(t)})}},$$

przy warunku

$$\|\mathbf{u}(s)\| = \|\mathbf{v}(t)\| = 1.$$

Ale

$$K_{u(s)} = k_u(\alpha_i, \alpha_j) = K_u$$

i

$$K_{v(t)} = k_v(\beta_i, \beta_j) = K_v.$$

Zatem właściwy wybór funkcji wagowych $\mathbf{u}(s)$ i $\mathbf{v}(t)$ jest równoważny wyborowi wektorów \mathbf{u} i \mathbf{v} tak, aby współczynnik

$$\hat{\rho}(\mathbf{u}, \mathbf{v}) = \frac{\text{tr}(\mathbf{K}_u \tilde{\mathbf{K}}_v)}{\sqrt{\text{tr}(\mathbf{K}_u \tilde{\mathbf{K}}_u) \text{tr}(\mathbf{K}_v \tilde{\mathbf{K}}_v)}}$$

osiągał maksimum przy warunku

$$\|\mathbf{u}\| = \|\mathbf{v}\| = 1.$$

W celu maksymalizacji powyższego współczynnika możemy użyć np. wyników Changa i innych (2013).

Współczynniki rzutu i tej realizacji $\mathbf{x}_i(s)$ procesu \mathbf{X}_s na k tą funkcjonalną zmienną kanoniczną są równe

$$\hat{U}_{ik} = \langle \hat{\mathbf{u}}_k(s), \mathbf{x}_i(s) \rangle = \int_{I_1} \hat{\mathbf{u}}_k^T(s) \mathbf{x}_i(s) dt = \hat{\boldsymbol{\alpha}}_i^T \hat{\mathbf{u}}_k.$$

Analogicznie współczynniki rzutu i tej realizacji $\mathbf{y}_i(t)$ procesu \mathbf{Y}_t na k tą funkcjonalną zmienną kanoniczną są równe

$$\hat{V}_{ik} = \hat{\boldsymbol{\beta}}_i^T \hat{\mathbf{v}}_k,$$

gdzie $i = 1, 2, \dots, n$, $k = 1, \dots, \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$, gdzie $\mathbf{A} \in \mathbb{R}^{n \times p}$ i $\mathbf{B} \in \mathbb{R}^{n \times q}$, gdzie i te wiersze to odpowiednio $\hat{\boldsymbol{\alpha}}_i$ i $\hat{\boldsymbol{\beta}}_i$.

Jako przykład posłużyły nam dane Światowego Forum Ekonomicznego (<http://www.weforum.org>) dotyczące wskaźników socjo-ekonomicznych 38 krajów w latach 2008-2015.

Kraje użyte w analizie.

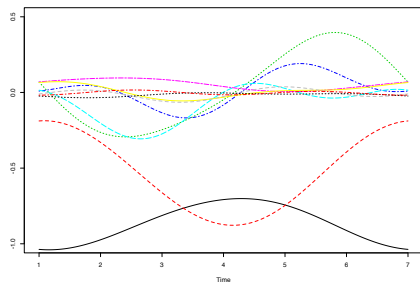
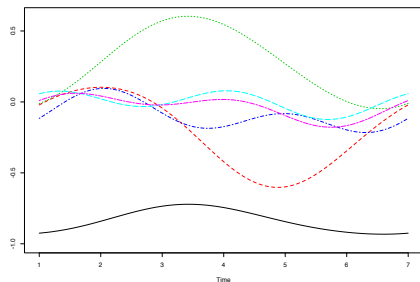
1	Albania (AL)	14	Greece (GR)	27	Poland (PL)
2	Austria (AT)	15	Hungary (HU)	28	Portugal (PT)
3	Belgium (BE)	16	Iceland (IS)	29	Romania (RO)
4	Bosnia and Herzegovina (BA)	17	Ireland (IE)	30	Russian Federation (RU)
5	Bulgaria (BG)	18	Italy (IT)	31	Serbia (XS)
6	Croatia (HR)	19	Latvia (LV)	32	Slovak Republic (SK)
7	Cyprus (CY)	20	Lithuania(LT)	33	Slovenia (SI)
8	Czech Republic (CZ)	21	Luxembourg (LU)	34	Spain (ES)
9	Denmark (DK)	22	Macedonia FYR (MK)	35	Sweden (SE)
10	Estonia (EE)	23	Malta (MT)	36	Switzerland (CH)
11	Finland (FI)	24	Montenegro (ME)	37	Ukraine (UA)
12	France (FR)	25	Netherlands (NL)	38	United Kingdom (GB)
13	Germany (DE)	26	Norway (NO)		

Zmienne użyte w analizie.

X (Higher education and training)	Y (Goods market efficiency)
1. Quality of the educational system	1. Intensity of local competition
2. Quality of math and science education	2. Extent of market dominance
3. Quality of management schools	3. Effectiveness of anti-monopoly policy
4. Internet access in schools	4. Agricultural policy costs
5. Local availability of specialized research and training services	5. Prevalence of trade barriers
6. Extent of staff training	6. Prevalence of foreign ownership
	7. Business impact of rules on FDI
	8. Burden of customs procedures
	9. Degree of customer orientation
	10. Buyer sophistication

Przykład

W pierwszym kroku wygładziliśmy dane. Została użyta baza Fouriera, dla każdej zmiennej ustaliliśmy maksymalną wielkość bazy na 5.



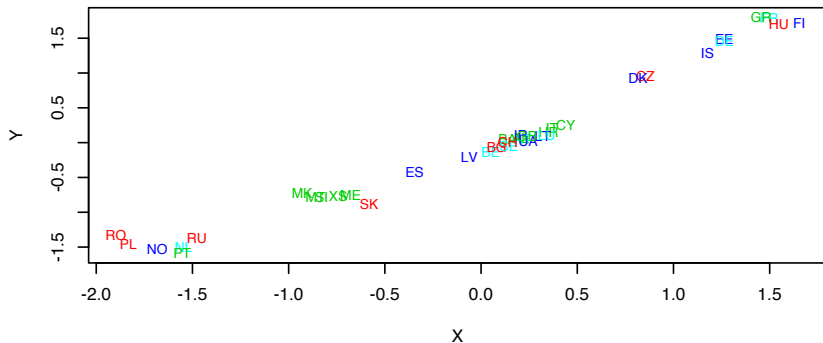
Funkcje wagowe dla pierwszej zmiennej kanonicznej U_1 (po lewej) i V_1 (po prawej).

Pola pod wartością bezwzględną funkcji wagowych.

Zmienna oryginalna	Pole	Udział (w %)
Pierwsza funkcjonalna składowa główna (X)		
1 (Quality of the educational system)	5.008	51.7
2 (Quality of math and science education)	1.724	17.8
5 (Local availability of specialized research and training services)	1.567	16.2
3	0.713	7.4
6	0.351	3.6
4	0.317	3.3
Pierwsza funkcjonalna składowa główna (Y)		
4 (Agricultural policy costs)	5.187	44.8
1 (Intensity of local competition)	3.194	27.6
5 (Prevalence of trade barriers)	1.287	11.1
10	0.580	5.0
2	0.511	4.4
7	0.323	2.8
9	0.206	1.8
8	0.152	1.3
3	0.091	0.8
6	0.057	0.5








Przykład

Względne położenie krajów w układzie (\hat{U}_1, \hat{V}_1) funkcjonalnych zmiennych kanonicznych (KTA) możemy zaobserwować poniżej.



Rzut krajów na płaszczyznę (\hat{U}_1, \hat{V}_1) .

Kraje z największą wartością funkcjonalnych zmiennych kanonicznych U_1 i V_1 to: Finlandia (FI), Francja (FR), Węgry (HU), Grecja (GR), Estonia (EE), Niemcy (DE), Islandia (IS), Czechy (CZ) i Dania (DK). Kraje z najmniejszą wartością funkcjonalnych zmiennych kanonicznych U_1 i V_1 to: Rumunia (RO), Polska (PL), Norwegia (NO), Portugalia (PT), Holandia (NL) i Rosja (RU). Pozostałe kraje tworzą grupę pośrednią.

-  CHANG, B., KRUGER, U., KUSTRA, R., ZHANG, J., (2013): Canonical Correlation Analysis based on Hilbert-Schmidt Independence Criterion and Centered Kernel Target Alignment. *Proceedings of the 30th International Conference on Machine Learning* 28(2):316–324.
-  CORTES, C., MOHRI, M., ROSTAMIZADEH, A., (2012): Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research* 13:795–828.
-  GÓRECKI, T., KRZYŚKO, M., WASZAK, Ł. and WOŁYŃSKI, W. (2014): Methods of reducing dimension for functional data. *Statistics in Transition new series* 15, 231–242.
-  GÓRECKI, T., KRZYŚKO, M., WOŁYŃSKI, W. (2017): Correlation analysis for multivariate functional data. *Studies in Classification, Data Analysis, and Knowledge Organization: Data Science* 243–258.
-  GÓRECKI, T., ŁUCZAK, M. (2015): Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications* 42(5), 2305–2312.
-  KONG, J., WANG, S., WAHBA G. (2015): Using distance covariance for improved variable selection with application to learning genetic risk models. *Statistics in Medicine* 34(10), 1708–1720.
-  LICHMAN, M. (2013): UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science.



RAMSAY, J.O., SILVERMAN, B.W. (2005): *Functional Data Analysis*. Springer, New York.



SZÉKELY, G.J., RIZZO, M.L., BAKIROV, N.K. (2007): Measuring and testing dependence by correlation of distances. *The Annals of Statistics* 35(6), 2769–2794.