

*Jądrowa i funkcjonalna analiza składowych głównych**

Mirosław Krzyśko

Wydział Matematyki i Informatyki UAM Poznań

* Referat ten jest przygotowany na podstawie wspólnych wyników uzyskanych z Karolem Deręgowskim z Wydziału Zarządzania Państwowej Wyższej Szkoły Zawodowej w Kaliszu oraz z Tomaszem Góreckim z Wydziału Matematyki i Informatyki UAM

Oznaczenia

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu},$$

$$\text{Var}(\mathbf{X}) = E\{[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]'\} = \boldsymbol{\Sigma} = (\sigma_{ij}),$$

gdzie

$$\sigma_{ij} = \text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j), \quad i, j = 1, 2, \dots, p.$$

Własności macierzy kowariancji

- Σ jest macierzą symetryczną, ponieważ $\sigma_{ij} = \sigma_{ji}$.
- $\sigma_{ii} = \sigma_i^2 = \text{Var}(X_i)$.
- Σ jest macierzą nieujemnie określoną, tj.

$$\forall \mathbf{t} \in \mathbb{R}^p, \mathbf{t}'\Sigma\mathbf{t} \geq 0.$$

Wielowymiarowy rozkład normalny

Mówimy, że $X \sim N(\mu, \sigma^2)$, gdy

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = \\ &= (2\pi)^{-\frac{1}{2}} (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right], \\ &x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0. \end{aligned}$$

Mówimy, że $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, gdy

$$\begin{aligned} f(\mathbf{x}) &= (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right], \\ &\mathbf{x} \in \mathbb{R}^p, \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} > 0. \end{aligned}$$

Próba losowa

Ciąg niezależnych wektorów losowych $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ o identycznym rozkładzie prawdopodobieństwa nazywamy próbą losową z tego rozkładu. Próbę losową zapisać możemy w postaci macierzy losowej

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$$

rozmiaru $p \times n$.

Statystyki z próby

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i = \text{wektor średnich}$$

$$\mathbf{A} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' = \text{macierz sum kwadratów i iloczynów.}$$

I_n – macierz jednostkowa stopnia n ,

\mathbf{J} – macierz stopnia n złożona z samych jedynek,

$\mathbf{1}_n$ – wektor kolumnowy złożony z samych jedynek.

$$\bar{\mathbf{X}} = \frac{1}{n} \mathbf{X} \mathbf{1}_n,$$

$$\mathbf{A} = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i' - n \bar{\mathbf{X}} \bar{\mathbf{X}}' = \mathbf{X} \left(I_n - \frac{1}{n} \mathbf{J} \right) \mathbf{X}'.$$

Próba z rozkładu $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- Jeżeli $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ jest próbą losową z rozkładu $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, gdzie $\boldsymbol{\Sigma} > 0$, to macierz \mathbf{A} jest dodatnio określona z prawdopodobieństwem jeden wtedy i tylko wtedy, gdy $n > p$.
- Jeżeli $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ jest próbą losową z rozkładu $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, gdzie $\boldsymbol{\Sigma} > 0$ oraz $n > p$, to statystyka $\bar{\mathbf{X}}$ jest nieobciążonym estymatorem $\boldsymbol{\mu}$ oraz statystyka $\mathbf{S} = \frac{1}{n-1}\mathbf{A}$ jest nieobciążonym estymatorem $\boldsymbol{\Sigma}$, tj.

$$E(\bar{\mathbf{X}}) = \boldsymbol{\mu} \text{ oraz } E(\mathbf{S}) = \boldsymbol{\Sigma}.$$

Składowe główne klasyczne

Dany jest wektor losowy \mathbf{X} . Bierzemy pod uwagę kombinację liniową

$$Z = a_1X_1 + a_2X_2 + \cdots + a_pX_p = \mathbf{a}'\mathbf{X}.$$

Spośród wszystkich wektorów $\mathbf{a} \in \mathbb{R}^p$ chcemy znaleźć wektor \mathbf{a} taki, by jego długość wynosiła 1 oraz by wariancja zmiennej Z była maksymalna.

Składowe główne klasyczne

Mamy

$$\text{Var}(Z) = \text{Var}(\mathbf{a}'\mathbf{X}) = \mathbf{a}' \text{Var}(\mathbf{X})\mathbf{a} = \mathbf{a}'\Sigma\mathbf{a}.$$

Niech

$$\lambda_1 = \sup_{\mathbf{a} \in \mathbb{R}^p, \mathbf{a}'\mathbf{a}=1} \mathbf{a}'\Sigma\mathbf{a}.$$

Wektor \mathbf{a}_1 maksymalizujący wariancję $\mathbf{a}'\Sigma\mathbf{a}$ przy dodatkowym warunku $\mathbf{a}'_1\mathbf{a}_1 = 1$ jest wektorem własnym macierzy Σ odpowiadającym największej wartości własnej λ_1 tej macierzy lub inaczej największemu pierwiastkowi równania wyznacznikowego

$$|\Sigma - \lambda I_p| = 0.$$

Składowe główne klasyczne

Wektor \mathbf{a}_1 spełnia zatem równanie

$$(\Sigma - \lambda_1 I_p) \mathbf{a}_1 = \mathbf{0}.$$

Zmienną $Z_1 = \mathbf{a}'_1 \mathbf{X}$ nazywamy pierwszą składową główną wektora \mathbf{X} . W celu wyznaczenia drugiej składowej konstruujemy kombinację liniową $Z_2 = \mathbf{a}'_2 \mathbf{X}$ taką, by $\mathbf{a}'_1 \mathbf{a}_2 = 0$ oraz by jej wariancja była maksymalna przy dodatkowym warunku $\mathbf{a}'_2 \mathbf{a}_2 = 1$. Szukamy zatem

$$\lambda_2 = \sup_{\mathbf{a} \in \mathbb{R}^p, \mathbf{a}'_1 \mathbf{a} = 0, \mathbf{a}' \mathbf{a} = 1} \mathbf{a}' \Sigma \mathbf{a}.$$

Składowe główne klasyczne

Współczynnik korelacji między składowymi głównymi $Z_1 = \mathbf{a}'_1 \mathbf{X}$ i $Z_2 = \mathbf{a}'_2 \mathbf{X}$ jest równy

$$r_{12} = \frac{\mathbf{a}'_1 \Sigma \mathbf{a}_2}{\sqrt{\lambda_1 \lambda_2}}.$$

Ponieważ $\Sigma \mathbf{a}_2 = \lambda_2 \mathbf{a}_2$, to $\mathbf{a}'_1 \Sigma \mathbf{a}_2 = \lambda_2 \mathbf{a}'_1 \mathbf{a}_2 = 0$, bo $\mathbf{a}'_1 \mathbf{a}_2 = 0$.

Zatem $r_{12} = 0$ i zmienne Z_1 i Z_2 są nieskorelowane. Kontynuując ten proces otrzymamy p składowych głównych

$Z_1 = \mathbf{a}'_1 \mathbf{X}$, $Z_2 = \mathbf{a}'_2 \mathbf{X}$, ..., $Z_p = \mathbf{a}'_p \mathbf{X}$ wzajemnie nieskorelowanych o wariancjach

$$\lambda_1 \geq \lambda_2 \dots \geq \lambda_p.$$

Składowe główne z próby

W praktyce macierz Σ nie jest znana i musimy zastąpić ją estymatorem z próby \mathbf{S} . Zauważmy, że

$$\text{tr } \mathbf{S} = \sum_{i=1}^p \text{Var}(X_i) = \sum_{i=1}^p \lambda_i,$$

czyli suma wariancji zmiennych pierwotnych jest równa sumie wariancji składowych głównych. Zatem wskaźnik

$$\frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_p} 100\%$$

jest miarą wyjaśniania całkowitej zmienności składowych wektora \mathbf{X} przez pierwszych k składowych głównych.

- Niech \mathbf{X} będzie macierzą rozmiaru $n \times p$ złożoną z n p -wymiarowych obserwacji $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^p$, tj.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \dots \\ \mathbf{x}'_n \end{bmatrix}.$$

- Każdą obserwację $\mathbf{x}_i \in \mathbb{R}^p$ przekształcamy za pomocą nieliniowej funkcji

$$\Phi : \mathbb{R}^p \rightarrow H,$$

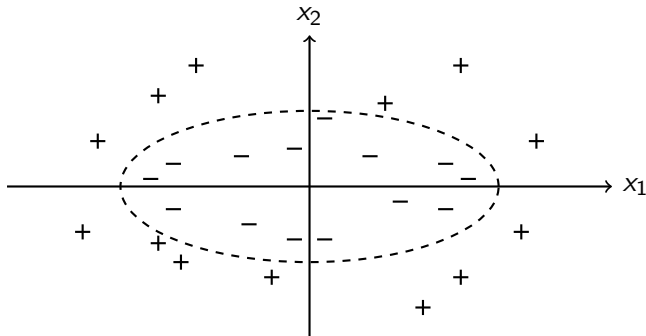
gdzie H ma wymiar $N_H > p$.

Rozpatrzmy następujący przykład.

Niech $\mathbf{x} = (x_1, x_2)'$. Definiujemy $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ jako

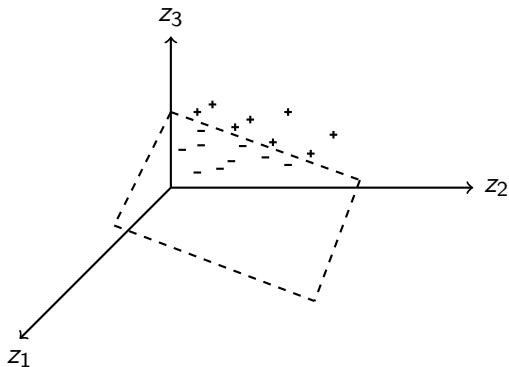
$$\Phi(\mathbf{x}) = \Phi(x_1, x_2) = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}.$$

Przypuśćmy, że oryginalne dane przedstawia poniższy rysunek.



Widzimy, że oryginalne dane na płaszczyźnie nie mogą być rozdzielone liniowo.

Rysunek poniżej przedstawia te same dane po przekształceniu za pomocą funkcji Φ .



Widzimy, że przekształcenie Φ powoduje liniową rozdzielczość danych w przestrzeni \mathbb{R}^3 .

- W naszym przykładzie znana była postać funkcji Φ . Okazuje się, że pod pewnymi warunkami, wystarczy tylko znajomość iloczynu skalarnego funkcji Φ .
- Weźmy pod uwagę iloczyn skalarny

$$\begin{aligned} \Phi'(\mathbf{x})\Phi(\mathbf{y}) &= \left[x_1^2, \sqrt{2}x_1x_2, x_2^2 \right] \begin{bmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \end{bmatrix} = \\ &= (x_1y_1 + x_2y_2)^2 = (\mathbf{x}'\mathbf{y})^2 = k(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Widzimy, że iloczyn skalarny funkcji Φ w przestrzeni \mathbb{R}^3 jest określony przez pewną funkcję $k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$.

Prawdziwe jest następujące twierdzenie.

Twierdzenie (Moore'a-Aronszajna (Aronszajn, 1950))

Niech

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

będzie rzeczywistą, symetryczną funkcją ciągłą i niech H będzie przestrzenią Hilberta. Wówczas istnieje przekształcenie

$$\Phi : \mathbb{R}^p \rightarrow H$$

takie, że

$$k(\mathbf{x}, \mathbf{y}) = \Phi'(\mathbf{x})\Phi(\mathbf{y}),$$

wtedy i tylko wtedy, gdy macierz $\mathbf{K} = (k_{ij})$, gdzie $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, 2, \dots, n$, jest macierzą nieujemnie określoną.

- Twierdzenie to mówi, że w przestrzeni Hilberta istnieje wzajemnie jednoznaczna odpowiedniość między funkcjami Φ a ich iloczynami skalarnymi .
- Macierz \mathbf{K} nazywamy macierzą jądrową, a funkcję $k(\mathbf{x}, \mathbf{y})$ nazywamy funkcją jądrową lub krótko jądrem.
- Piękno przekształcenia Φ polega na tym, że w każdym algorytmie, w którym dane występują tylko w postaci iloczynów skalarnych, nie musimy explicite znać postaci funkcji $\Phi(\cdot)$, lecz tylko wybrać postać jądra $k(\mathbf{x}, \mathbf{y})$.
- Stwierdzenie, czy funkcja $k(\mathbf{x}, \mathbf{y})$ jest jądrem może być sprawą trudną. W praktyce, istnieje wiele funkcji, o których wiadomo z literatury, że są jądrami.

Tabela poniżej zawiera przykłady funkcji $k(\mathbf{x}, \mathbf{y})$, które są jądrami.

Jądro	$k(\mathbf{x}, \mathbf{y})$
Jednorodne jądro wielomianowe	$(\mathbf{x}'\mathbf{y})^d$
Niejednorodne jądro wielomianowe	$(\mathbf{x}'\mathbf{y} + c)^d, c > 0$
Laplace'a	$\exp(-\sigma\ \mathbf{x} - \mathbf{y}\), \sigma > 0$
Gausa	$\exp(-\sigma\ \mathbf{x} - \mathbf{y}\ ^2), \sigma > 0$

Zauważmy, że w naszym przykładzie jądro $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}'\mathbf{y})^2$ było jednorodnym jądrem wielomianowym stopnia $d = 2$.

Pomysł wykorzystania funkcji jądrowych do konstrukcji nieliniowych składowych głównych pochodzi od Schölkopfa, Smoli i Mülera (1998) i sprowadza się do tego, że w pierworyginalne dane przekształcane są za pomocą pewnej nieliniowej funkcji Φ do przestrzeni Hilberta z jądrem spełniającym warunki tw. Mercera (Mercer (1909); patrz Cristianini and Shawe-Taylor (2000) oraz Vapnik (1998)), a następnie w tej nowej przestrzeni, w sposób tradycyjny, konstruowane są liniowe składowe główne. Posługiwanie się jądrami spełniającymi założenia tw. Mercera nastręcza pewne trudności. O ile takie jądra spełniają warunek $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$, to stwierdzenie odwrotne nie jest prawdziwe. Ponieważ warunek ten jest dla nas bardzo istotny, to właściwą klasą jąder, które powinniśmy brać pod uwagę jest klasa jąder nieujemnie określonych. W dalszych rozważaniach wybierzemy klasę funkcji Φ z przestrzeni Hilberta z nieujemnie określonym jądrem, ponieważ wówczas tw. Moore'a-Aronszajna gwarantuje nam wzajemnie jednoznaczną odpowiedniość między funkcjami Φ a ich iloczynami skalarnymi.

W celu konstruowania liniowych składowych głównych w przestrzeni H znajdujemy wartości własne $\lambda \geq 0$ i niezerowe wektory własne $\mathbf{v} \in H$ macierzy kowariancji

$$\mathbf{C} = \frac{1}{n} \sum_{j=1}^n \Phi(\mathbf{x}_j) \Phi'(\mathbf{x}_j) \quad (1)$$

nieliniowo przekształconych za pomocą funkcji Φ wektorów $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, a następnie scentrowanych.

Jeżeli \mathbf{v} jest wektorem własnym macierzy \mathbf{C} odpowiadającym wartości własnej $\lambda \geq 0$, to

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \quad (2)$$

lub równoważnie

$$\Phi'(\mathbf{x}_j)\mathbf{C}\mathbf{v} = \lambda\Phi'(\mathbf{x}_j)\mathbf{v}, \quad j = 1, 2, \dots, n. \quad (3)$$

Ponieważ

$$\mathbf{C}\mathbf{v} = \frac{1}{n} \sum_{j=1}^n \Phi(\mathbf{x}_j)\Phi'(\mathbf{x}_j)\mathbf{v}, \quad (4)$$

to wszystkie rozwiązania \mathbf{v} odpowiadające niezerowym wartościom własnym λ należą do podprzestrzeni rozpiętej przez wektory $\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_n)$.

Stąd istnieją współczynniki α_j , $j = 1, 2, \dots, n$, takie, że

$$\mathbf{v} = \sum_{j=1}^n \alpha_j \Phi(\mathbf{x}_j). \quad (5)$$

Podstawiając (5) do (3) dostajemy:

$$\begin{aligned} & \frac{1}{n} \sum_{j=1}^n \alpha_j \Phi(\mathbf{x}_j) \sum_{k=1}^n \Phi(\mathbf{x}_k) \Phi'(\mathbf{x}_k) \Phi(\mathbf{x}_j) = \\ & = \lambda \sum_{k=1}^n \alpha_k \Phi'(\mathbf{x}_i) \Phi(\mathbf{x}_k), \quad i = 1, 2, \dots, n. \end{aligned} \quad (6)$$

W zapisie macierzowym równość (6) przybiera postać

$$\mathbf{K}^2 \boldsymbol{\alpha} = n\lambda \mathbf{K} \boldsymbol{\alpha}, \quad (7)$$

gdzie $\mathbf{K} = (k_{ij})$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)'$ lub

$$\mathbf{K}^2 \boldsymbol{\alpha} = \tilde{\lambda} \mathbf{K} \boldsymbol{\alpha}, \quad (8)$$

gdzie $\tilde{\lambda} = n\lambda$.

Zauważmy, że każdy wektor $\alpha \neq \mathbf{0}$ będący rozwiązaniem równania

$$K\alpha = \tilde{\lambda}\alpha \quad (9)$$

jest zarazem rozwiązaniem równania (8), i że rozwiązania równań (8) i (9) różnią się tylko wektorami własnymi macierzy K odpowiadającymi zerowym wartościom własnym, co nie ma znaczenia w problemie składowych głównych.

W praktyce wektorów $\{\Phi(x_i)\}$, $i = 1, 2, \dots, n$ nie możemy scentrować ponieważ nie znamy funkcji Φ . Niech

$$\tilde{\Phi}_i = \Phi_i - \frac{1}{n} \sum_{k=1}^n \Phi_k$$

oraz

$$\tilde{K} = (\tilde{k}_{ij}) = (\langle \tilde{\Phi}_i, \tilde{\Phi}_j \rangle).$$

Wówczas

$$\tilde{K} = PKP,$$

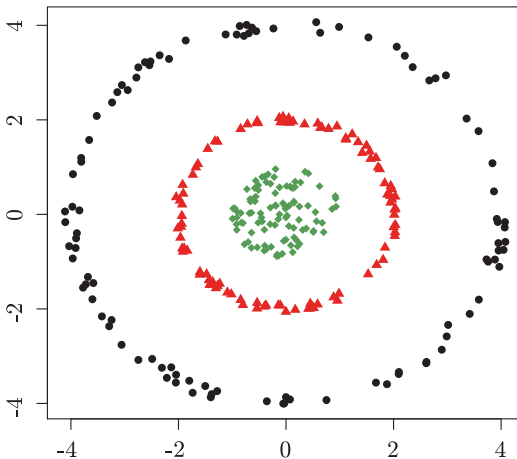
gdzie $P = (\delta_{ij} - 1/n)$ oraz K jest macierzą jądrową.

Przykład

Wygenerowano 100 punktów z rozkładu jednostajnego na kole jednostkowym, 100 punktów z rozkładu jednostajnego na okręgu o promieniu 2 oraz 100 punktów z rozkładu jednostajnego na okręgu o promieniu 4. W dwóch ostatnich przypadkach rozkłady te zostały zaburzone rozkładem $N(0, \sigma^2)$, gdzie $\sigma = 0.25$.

Przykład

Dane te pokazane są na poniższym rysunku.



Przykład

Dla tego typu danych sferycznych nie można skonstruować sensownie tradycyjnych składowych głównych. Z połączonych 300 obserwacji skonstruowane zostały składowe główne w oparciu o funkcję jądrową. Jako funkcję jądrową wybrano funkcję gaussowską postaci

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

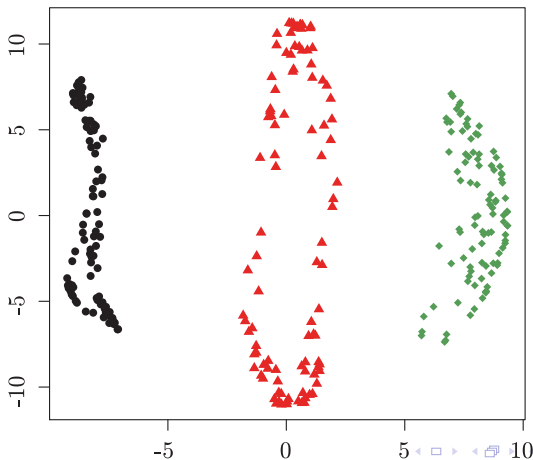
Utworzona została jądrowa macierz $\tilde{\mathbf{K}} = (\tilde{k}_{ij})$, gdzie $\tilde{k}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, 2, \dots, 300$, a następnie rozwiązano zagadnienie własne

$$\tilde{\mathbf{K}}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}.$$

Ponieważ funkcja jądrowa $k(\mathbf{x}, \mathbf{y})$ zależy od parametru $\sigma > 0$, to wszystkie wielkości $\tilde{\mathbf{K}}$, $\boldsymbol{\alpha}$ oraz λ zależą od tego parametru.

Przykład

Trzysta wygenerowanych punktów w układzie dwóch pierwszych składowych głównych pokazanych jest na poniższym rysunku.



Wiele danych finansowych, meteorologicznych, itp., notowanych jest w dyskretnych punktach czasowych. Niech y_{ij} oznacza zaobserwowaną wartość badanej cechy statystycznej na i -tej jednostce w j -tym momencie czasowym, gdzie $i = 1, 2, \dots, N$, $j = 1, 2, \dots, J_i$. Momenty obserwacji danej cechy statystycznej od jednostki do jednostki mogą się zmieniać, a odstępy między momentami czasowymi nie muszą być jednakowe. Wówczas nasze dane składają się z par $\{t_{ij}, y_{ij}\}$, gdzie $t_{ij} \in [0, T]$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, J_i$.

Jednakże w wielu przypadkach wygodniej jest posługiwać się ciągłymi funkcjami czasu $x(t)$, $t \in [0, T]$. W tym przypadku mówi się o **danych funkcjonalnych** (Ramsay i Danzell (1991)). Zazwyczaj dane funkcjonalne składają się z N niezależnych realizacji

$$\{x_i(t), i = 1, 2, \dots, N, t \in [0, T]\},$$

pewnego procesu losowego $X(t)$.

W takiej sytuacji możemy dokonać konwersji danych dyskretnych $\{t_{ij}, y_{ij}\}$ do danych funkcjonalnych $\{x_i(t), i = 1, 2, \dots, N, t \in [0, T]\}$. Konwersja z dyskretnych danych do funkcji ciągłych wymaga wygładzania (Ramsay i Silverman (2005)). Proces konwersji przeprowadzamy dla każdego $i \in \{1, 2, \dots, N\}$ oddzielnie. Stąd nasze dalsze rozważania dotyczyć będą pojedynczej funkcji $x(t)$, gdzie $t \in [0, T]$.

Jednym ze sposobów wygładzania jest przedstawienie funkcji $x(t)$ jako kombinacji liniowej $K + 1$ ortonormalnych funkcji bazowych φ_k :

$$x(t) = \sum_{k=0}^K c_k \varphi_k(t), \quad t \in [0, T].$$

Współczynniki c_k tego rozwinięcia dobiera się metodą najmniejszych kwadratów, tj. tak aby minimalizowały funkcję:

$$S(c) = (y - \Phi c)'(y - \Phi c),$$

gdzie $c = (c_0, c_1, \dots, c_K)'$ oraz Φ jest macierzą rozmiaru $J \times (K + 1)$ zawierającą wartości $\varphi_k(t_j)$. Różniczkując funkcję $S(c)$ względem wektora c otrzymujemy:

$$\hat{c} = (\Phi' \Phi)^{-1} \Phi' y.$$

Stopień gładkości zależy od K , ponieważ małe (duże) wartości K powodują większe (mniejsze) wygładzanie krzywych. Optymalna wartość K wybrana jest dla każdej funkcji $x(t)$ za pomocą **bayesowskiego kryterium informacyjnego (BIC)** lub **kryterium informacyjnego AKAIKE (AIC)**, a następnie z wartości K odpowiadających wszystkim funkcjom wybierana jest wartość modalna, jako wspólna wartość dla wszystkich rozwinięć funkcji $x(t)$. Kryterium BIC mierzy dokładność dopasowania modelu, natomiast AIC mierzy jakość predykcji modelu (Shmueli (2010)).

Konstrukcja funkcjonalnych składowych głównych

Założmy, że obserwujemy realizację procesu stochastycznego $X(t) \in L_2([0, T])$, gdzie $L_2([0, T])$ jest przestrzenią HILBERTA funkcji całkownych z kwadratem na przedziale $[0, T]$ wyposażoną w iloczyn skalarny

$$\langle u, v \rangle = \int u(s)v(s)ds.$$

Założmy ponadto, że $EX(t) = 0$ oraz

$$\int E\|X\|^2 = E[\langle X, X \rangle] = E \int X^2(s)ds < \infty.$$

Składowe główne możemy zdefiniować zarówno dla wektorów losowych skończenie wymiarowych $X \in \mathbb{R}^k$, jak i również dla procesów stochastycznych $X(t) \in L_2([0, T])$ w sposób następujący. Niech $H = \mathbb{R}^p$ w przypadku wektorowym oraz $H = L_2([0, T])$ w przypadku funkcjonalnym.

Konstrukcja funkcjonalnych składowych głównych

Wówczas pierwsza wartość własna λ_1 i związana z nią **funkcja wagowa** lub wektor wag u_1 są zdefiniowane następująco:

$$\lambda_1 = \sup_{u \in H} \text{Var}(\langle u, X \rangle) = \text{Var}(\langle u_1, X \rangle),$$

gdzie wektor u_1 podlega ograniczeniu postaci

$$\|u_1\| = 1. \tag{10}$$

Ograniczenie to zapewnia jednoznaczność pierwszej składowej głównej z dokładnością do znaku.

Konstrukcja funkcjonalnych składowych głównych

Postępując analogicznie, k -tą wartość własną λ_k oraz odpowiadający jej wektor wag lub funkcję wagową u_k definiuje się następująco:

$$\lambda_k = \sup_{u \in H} \text{Var}(\langle u, X \rangle) = \text{Var}(\langle u_k, X \rangle),$$

gdzie wektor u_k podlega warunkowi $\|u_k\| = 1$ oraz k -ta składowa główna

$$U_k = \langle u_k, X \rangle$$

jest nieskorelowana z $(k - 1)$ pierwszymi składowymi głównymi $\{U_i, i = 1, \dots, k - 1\}$.

Parę (λ_k, u_k) będziemy nazywać k -tą **konfiguracją główną**.

Konstrukcja funkcjonalnych składowych głównych

W dalszym ciągu będziemy zakładać, że proces stochastyczny $X(t)$ może być reprezentowany za pomocą skończonej liczby ortonormalnych funkcji bazowych, tj., że $X(t)$ można zapisać w postaci:

$$X(t) = \sum_{k=0}^K c_k \varphi_k(t), \quad t \in [0, T],$$

gdzie $\{\varphi_k\}$ są $(K + 1)$ pierwszymi elementami bazy ortonormalnej $L_2([0, T])$, a $\{c_k\}$ są zmiennymi losowymi z zerowymi wartościami oczekiwanymi oraz skończonymi wariancjami. Niech:

$$\begin{aligned} \varphi(t) &= (\varphi_0(t), \varphi_1(t), \dots, \varphi_K(t))', \\ c &= (c_0, c_1, \dots, c_K)', \quad 0 < K < \infty, \end{aligned}$$

gdzie $E(c) = 0$ oraz $\text{Var}(c) = \Sigma > 0$. Wówczas proces $X(t)$ można zapisać w postaci:

$$X(t) = c' \varphi(t), \quad t \in [0, T].$$

Konstrukcja funkcjonalnych składowych głównych

Twierdzenie

k -ta konfiguracja główna wektora losowego c zdefiniowana przez (σ_k, u_k) , jest związana z k -tą konfiguracją główną procesu stochastycznego $X(t)$, $(\lambda_k, u_k(t))$ w następujący sposób:

$$u_k(t) = u_k' \varphi(t), \quad \lambda_k = \sigma_k.$$

Widzimy, że analiza składowych głównych dla procesów $X(t)$ skończenie wymiarowych jest równoważna wielowymiarowej analizie składowych głównych dla wektorów losowych $c = (c_0, c_1, \dots, c_K)'$.

Konstrukcja funkcjonalnych składowych głównych

Analiza składowych głównych dla wektorów losowych c bazuje na macierzy kowariancji Σ . W praktyce macierz ta nie jest znana. Oceniamy ją na podstawie N niezależnych realizacji wektora losowego c zestawionych w macierz postaci:

$$\hat{C} = \begin{bmatrix} \hat{c}_{10} & \hat{c}_{11} & \dots & \hat{c}_{1K} \\ \hat{c}_{20} & \hat{c}_{21} & \dots & \hat{c}_{2K} \\ \dots & \dots & \dots & \dots \\ \hat{c}_{N0} & \hat{c}_{N1} & \dots & \hat{c}_{NK} \end{bmatrix}$$

gdzie \hat{c}_{ik} są estymatorami uzyskanymi **metodą najmniejszych kwadratów** parametrów c_{ik} w reprezentacji

$$x_i(t) = \sum_{k=0}^K c_{ik} \varphi_k(t)$$

procesu $X(t)$.

Konstrukcja funkcjonalnych składowych głównych

Wówczas estymator $\hat{\Sigma}$ nieznaney macierzy Σ ma następującą postać:

$$\hat{\Sigma} = \frac{1}{N} \hat{C}' \hat{C}.$$

Jeżeli $N > K + 1$, to macierz $\hat{\Sigma}$ jest dodatnio określona z prawdopodobieństwem jeden. Dodatnia określoność tej macierzy jest pożądana, ponieważ tylko wówczas jej wartości własne i wektory własne są ocenami największej wiarygodności wartości własnych i wektorów własnych macierzy Σ i stąd składowe główne z próby będą ocenami największej wiarygodności populacyjnych składowych głównych (Seber (1984), s.197).

Konstrukcja funkcjonalnych składowych głównych

Następnie znajdujemy wartości własne $\hat{\lambda}_k$ i odpowiadające im wektory własne \hat{u}_k macierzy $\hat{\Sigma}$. Wszystkie wartości własne są rzeczywiste, dodatnie i różne. Mając wyznaczone wektory własne \hat{u}_k wyznaczamy funkcje własne (funkcje wagowe):

$$\hat{u}_k(t) = \hat{u}'_k \varphi(t), \quad t \in [0, T].$$

Stąd współrzędna rzutu i -tej realizacji $x_i(t)$ procesu $X(t)$ na j -tą funkcjonalną składową główną jest równa:

$$\begin{aligned} \hat{U}_{ij} &= \langle \hat{u}_j(t), x_i(t) \rangle = \int \hat{u}_j(t) x_i(t) dt = \sum_{k=0}^K \sum_{l=0}^K \hat{c}_{il} \hat{u}_{jk} \int \varphi_k(t) \varphi_l(t) dt = \\ &= \sum_{k=0}^K \hat{c}_{ik} \hat{u}_{jk} = \hat{c}'_i \hat{u}_j, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots \end{aligned}$$

Konstrukcja funkcjonalnych składowych głównych

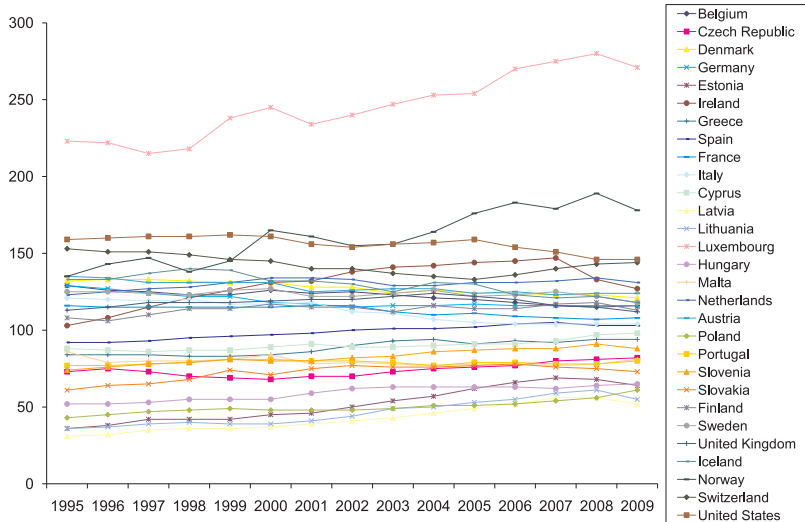
Uwaga

Oceny wektora losowego c można przekształcić do przestrzeni HILBERTA z nieujemnie określonym jądrem reprodukcującym i skonstruować jądrowe składowe główne dla danych funkcjonalnych.

Przykład

Wydajność opisanej metody została przetestowana na danych dotyczących PKB per capita (PPS) państw Unii Europejskiej (EU25=100) z lat 1995-2009.

Przykład



Przykład

Dyskretne szeregi czasowe zostały scentrowane a następnie przekształcone w funkcje ciągłe. Użyte zostały cztery różne bazy ortonormalne:

- 1 Baza FOURIERA w przestrzeni $L^2([0, T])$:
- 2 Baza sinusów w przestrzeni $L^2([0, T])$:
- 3 Baza cosinusów w przestrzeni $L^2([0, T])$:
- 4 Baza wielomianów LEGENDRE'A w przestrzeni $L^2([-1, 1])$:

Przykład

Optymalna liczba K funkcji bazowych $\varphi_k(t)$ została wybrana za pomocą kryterium informacyjnego AKAIKE (AIC) oraz bayesowskiego kryterium informacyjnego (BIC). Przyjeliśmy założenie, że parametr K nie może przyjmować wartości większej od 100.

FOURIER		Sinus		Cosinus		Wielomiany LEGENDRE'A	
AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
3	3	13	13	3	3	16	16

Przykład

Najczęściej badane obiekty są przestawiane jako punkty na wykresie w układzie dwóch pierwszych funkcjonalnych składowych głównych. W takim przypadku **kryterium jakości** skonstruowanych funkcjonalnych składowych głównych ma postać:

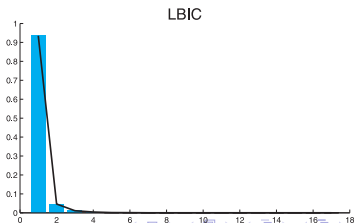
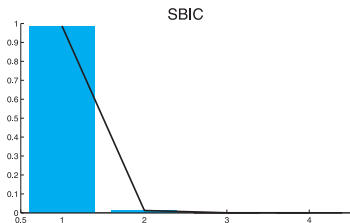
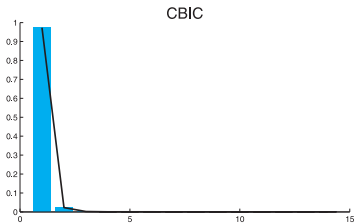
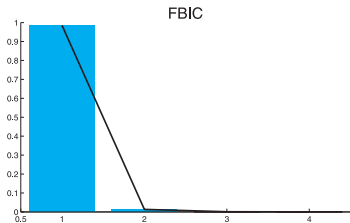
$$\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} 100\%,$$

gdzie $\lambda_1 \geq \lambda_2 \geq \dots$ są niezerowymi wartościami własnymi macierzy $\hat{\Sigma}$. Im większa wartość wyrażenia tym więcej zmienności zachowują dwie pierwsze funkcjonalne składowe główne.

FOURIER		Sinus		Cosinus		Wielomiany LEGENDRE'A	
AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
99.89	99.89	99.59	99.59	99.95	99.95	98.17	98.17

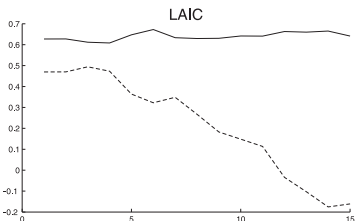
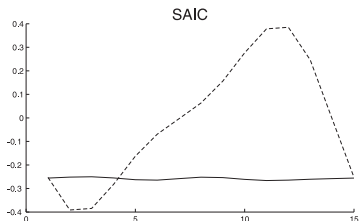
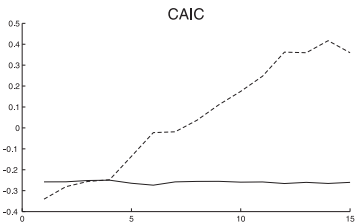
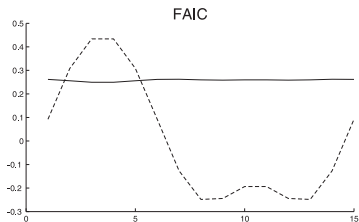
Przykład

Wykres osypiska dla wartości własnych.



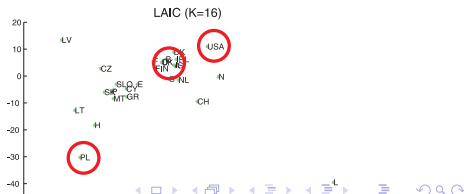
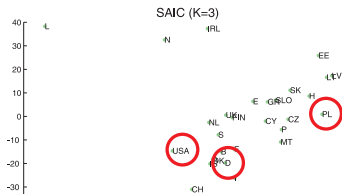
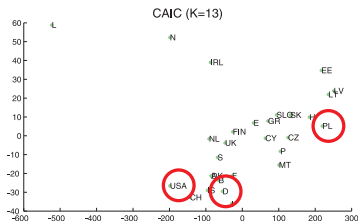
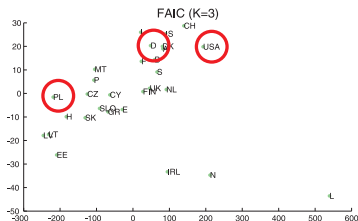
Przykład






Dwie pierwsze funkcje wagowe.



Przykład

Wykres rozrzutu w przestrzeni dwóch pierwszych funkcjonalnych składowych głównych.



-  Aronszajn N., Theory of reproducing kernels, Trans. Amer. Math. Soc. 68 (1950), 337–404.
-  Cristianini N., Shawe-Taylor J., An Introduction to Support Vector Machines , Cambridge University Press, 2000
-  Ramsay J.O., Dalzell C.J., Some tools for functional data analysis, J. Royal Statist. Soc. B 53 (1991), 539–572.
-  Ramsay J.O, Silverman B.W., Applied Functional Data Analysis, Springer, New York 2002.
-  Ramsay J.O, Silverman B.W., Functional Data Analysis, Second Edition, Springer, New York 2005.
-  Schölkopf B., Smola, A.J., Müller K.R., Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10, 1299–1319.
-  Seber G.A.F., Multivariate Observations, Wiley, New York 1984.
-  Shmueli G., To explain or to predict?, Statistical Science 25(3) (2010), 289–310.
-  Vapnik V., Statistical Learning Theory, Wiley, New York 1998.