

Zmienne dyskryminacyjne

Tomasz Górecki, Mirosław Krzyśko, Łukasz Waszak

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza

Wielowymiarowe metody statystyczne
Poznań 19.12.2012

Jeżeli mamy próby pochodzące z c grup, to często chcielibyśmy przedstawić je graficznie, by zobaczyć ich konfigurację lub też odrzucić obserwacje odstające. Jednakże taka prezentacja graficzna może być trudna nawet w przypadku obserwowania tylko trzech cech, a w przypadku większej liczby cech niemożliwa. Należy zatem poszukiwać innej możliwości prezentacji danych wielowymiarowych pochodzących z wielu grup.

Dla ułatwienia zadania, w pierwszym kroku, każdą p -wymiarową obserwację $\mathbf{x} = (x_1, x_2, \dots, x_p)'$ można przekształcić w obserwację jednowymiarową $u_1 = \mathbf{a}'_1 \mathbf{x} = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$, a otrzymane obserwacje jednowymiarowe przedstawić graficznie jako punkty na prostej. W drugim kroku możemy zdefiniować drugą kombinację liniową $u_2 = \mathbf{a}'_2 \mathbf{x} = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p$ nieskorelowaną z pierwszą i obserwacje przedstawić graficznie jako punkty płaszczyzny,

Ogólnie, celem jest skonstruowanie nowych zmiennych nieskorelowanych u_1, u_2, \dots, u_s , $s \leq p$, które będą kombinacjami liniowymi obserwacji pierwotnych x_1, x_2, \dots, x_p oraz w maksymalnym stopniu będą różnicowały (dyskryminowały) c grup, tj. w nowym układzie środki c grup będą maksymalnie rozsunięte, a obserwacje z danej grupy będą maksymalnie skupione wokół jej środka. Te nowe zmienne nazywają się zmiennymi dyskryminacyjnymi.

Dany jest zbiór n p -wymiarowych obserwacji $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, gdzie wektory $\mathbf{x}_i \in \mathbf{R}^p$ pogrupowane są w c rozłącznych klas i każdy wektor \mathbf{x}_i należy do jednej i tylko jednej klasy.

Niech $V = \{1, 2, \dots, n\}$ będzie zbiorem indeksów zbioru obserwacji $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ i niech zbiór V będzie podzielony na c rozłącznych podzbiorów V_i takich, że $V_i \cap V_j = \emptyset$ dla $i \neq j$, $\bigcup_{j=1}^c V_j = V$ oraz zbiór V_j zawiera n_j elementów: $\sum_{j=1}^c n_j = n$.

Niech

$$\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

będzie średnią z próby i niech

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{i \in V_j} \mathbf{x}_i$$

będzie średnią j -tej klasy, dla $j = 1, 2, \dots, c$. Oznaczmy przez \mathbf{S}_t macierz zmienności całkowitej oraz \mathbf{S}_b macierz zmienności międzyklasowej. Mamy

$$\mathbf{S}_t = \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})',$$

$$\mathbf{S}_b = \sum_{j=1}^c n_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})'.$$

Poszukujemy zbioru wektorów $\mathbf{a}_i \in \mathbf{R}^p$, które maksymalizują miarę różnicowania c klas

$$J(\mathbf{a}) = \frac{\mathbf{a}' \mathbf{S}_b \mathbf{a}}{\mathbf{a}' \mathbf{S}_t \mathbf{a}},$$

przy dodatkowym warunku ograniczającym

$$\mathbf{a}_i' \mathbf{S}_t \mathbf{a}_j = \delta_{ij} \text{ (delta Kroneckera)}$$

oznaczającym, że zmienne $u_i = \mathbf{a}_i' \mathbf{x}$, zwane zmiennymi dyskryminacyjnymi, są nieskorelowane i mają jednostkowe wariancje. Wektory \mathbf{a}_i nazywane są wektorami kierunkowymi.

Znalezienie wektorów kierunkowych sprowadza się do rozwiązania następującego uogólnionego zagadnienia własnego

$$\mathbf{S}_b \mathbf{a}_j = \lambda_j \mathbf{S}_t \mathbf{a}_j, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s > \lambda_{s+1} = 0, \quad (1)$$

gdzie $s = \min\{p, c - 1\}$, $c - 1 = \text{rzęd} \mathbf{S}_b$.

Ponieważ $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$, gdzie \mathbf{S}_w jest macierzą zmienności wewnątrzklasowej, to znalezienie klasycznych zmiennych dyskryminacyjnych jest równoważne znalezieniu rozwiązania uogólnionego zagadnienia własnego postaci

$$\mathbf{S}_b \mathbf{a} = \lambda / (1 - \lambda) \mathbf{S}_w \mathbf{a}.$$

Wektory własne są identyczne, a wartości własne proporcjonalne.

Jeżeli \mathbf{S}_t jest macierzą nieosobliwą, to otrzymujemy

$$\mathbf{S}_t^{-1} \mathbf{S}_b \mathbf{a} = \mathbf{a} \lambda.$$

Zatem pary $(\lambda_j, \mathbf{a}_j)$, gdzie λ_j jest wartością własną macierzy $\mathbf{S}_t^{-1} \mathbf{S}_b$, a \mathbf{a}_j jest odpowiadającym jej wektorem własnym, służą do konstrukcji zmiennych dyskryminacyjnych. Macierz $\mathbf{S}_t^{-1} \mathbf{S}_b$ nie jest macierzą symetryczną. Biorąc pod uwagę rozkład Cholesky'ego macierzy \mathbf{S}_t postaci $\mathbf{S}_t = \mathbf{U}' \mathbf{U}$ otrzymujemy $(\mathbf{U}^{-1})' \mathbf{S}_b \mathbf{U}^{-1} \mathbf{a} = \lambda \mathbf{a}$. Wartości własne macierzy symetrycznej $(\mathbf{U}^{-1})' \mathbf{S}_b \mathbf{U}^{-1}$ i macierzy $\mathbf{S}_t^{-1} \mathbf{S}_b$ są identyczne. Jeśli \mathbf{b} jest wektorem własnym macierzy $(\mathbf{U}^{-1})' \mathbf{S}_b \mathbf{U}^{-1}$, to $\mathbf{a} = \mathbf{U}^{-1} \mathbf{b}$ jest wektorem własnym macierzy $\mathbf{S}_t^{-1} \mathbf{S}_b$.

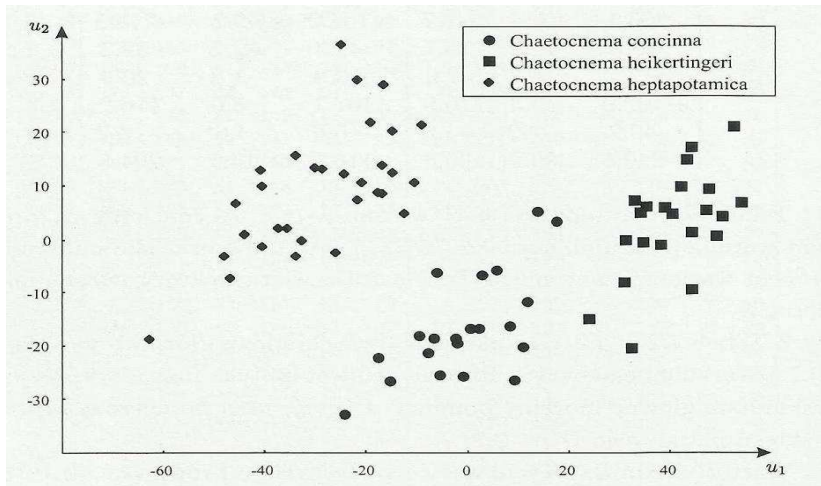
Przykład 1

Dane zostały zaczerpnięte z pracy Lubischewa (1962) i dotyczą pomiaru sześciu cech ($p = 6$) na okazach trzech męskich gatunków ($c = 3$) chrząszczy skaczących: 21 okazów gatunku ($n_1 = 21$) *Chaetocnema concinna*, 31 okazów gatunku ($n_2 = 31$) *Chaetocnema heikertingeri* oraz 22 okazy gatunku ($n_3 = 22$) *Chaetocnema heptapotamica* ($n = 74$).

Klasyczne zmienne dyskryminacyjne

Wartości własne: 967,7(76,7%); 226,6(18,0%)

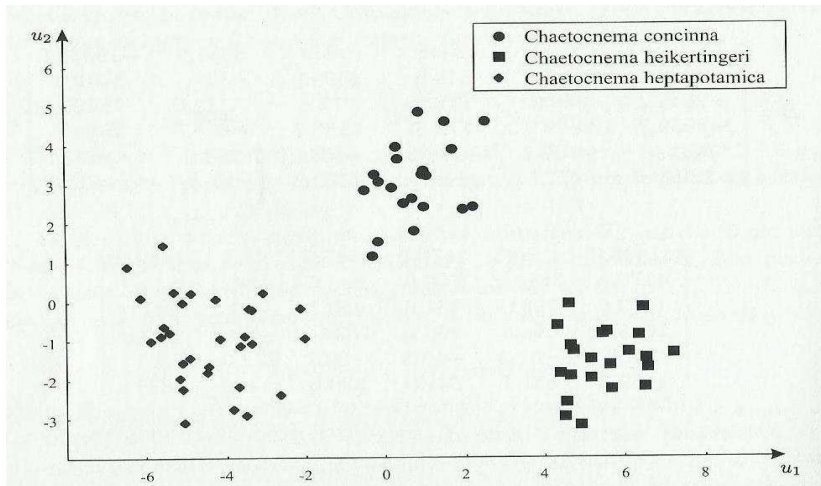
Rysunek: Rzut na przestrzeń 2 pierwszych składowych głównych



Klasyczne zmienne dyskryminacyjne

Wartości własne: 17,779(82%); 3,885(18%)

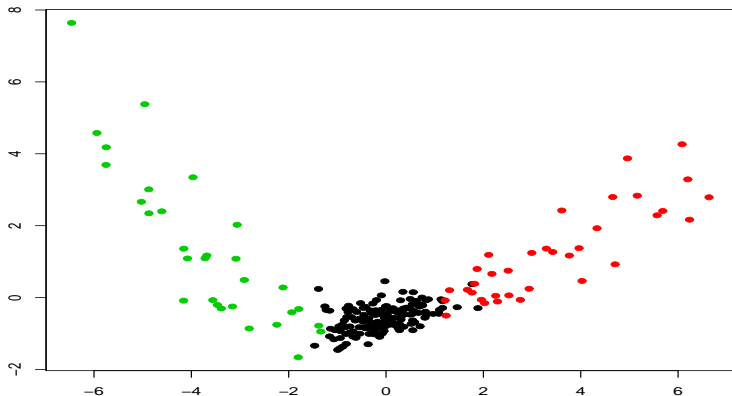
Rysunek: Rzut na przestrzeń 2 pierwszych zmiennych dyskryminacyjnych



Przykład 2

Zbiór danych Thyroid zawiera 215 obiektów ($n = 215$) opisanych za pomocą 5 cech ($p = 5$). Każdy obiekt należy do jednej z 3 klas ($c = 3$), a liczebności klas wynoszą odpowiednio 150, 35 i 30 ($n_1 = 150, n_2 = 35, n_3 = 30$).

Rysunek: Rzut na przestrzeń 2 pierwszych zmiennych dyskryminacyjnych



Jądrowe zmienne dyskryminacyjne wprowadzone zostały niezależnie przez Mikę i innych (1999) oraz Baudata i Anouara (2000). Metoda ta jest opisana w książce Shawe-Taylor and Cristianini (2004). Oryginalną przestrzeń \mathbf{R}^P przekształcamy nieliniowo w przestrzeń cech \mathcal{H}_k

$$\phi : \mathbf{R}^P \rightarrow \mathcal{H}_k,$$

gdzie ϕ funkcją wektorową, natomiast \mathcal{H}_k jest przestrzenią Hilberta z jądrem reprodukcującym.

Wektor $\phi(\mathbf{x}_i) = \tilde{\mathbf{x}}_i$ nazywa się wektorem cech odpowiadającym obserwacji $\mathbf{x}_i \in \mathbf{R}^p$, $i = 1, \dots, n$. Nieliniowe przekształcenie ϕ nie jest ogólnie znane, natomiast wybieramy znaną postać nieujemnie określonej funkcji jądrowej

$$k(\mathbf{x}, \mathbf{y}) = \phi'(\mathbf{x})\phi(\mathbf{y}) = \tilde{\mathbf{x}}'\tilde{\mathbf{y}}.$$

Do najbardziej popularnych funkcji jądrowych należą:

jądro gaussowskie $k(\mathbf{x}, \mathbf{y}) = \exp(-c\|\mathbf{x} - \mathbf{y}\|^2)$, $c > 0$

jądro wielomianowe $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}'\mathbf{y} + c)^d$, $c > 0$, $d > 0$.

Niech $\tilde{\mathbf{S}}_b$ i $\tilde{\mathbf{S}}_t$ oznacza odpowiednio macierz zmienności międzyklasowej oraz całkowitej w przestrzeni cech. Otrzymujemy

$$\tilde{\mathbf{S}}_b = \sum_{j=1}^c n_j (\tilde{\mathbf{m}}_j - \tilde{\mathbf{m}})(\tilde{\mathbf{m}}_j - \tilde{\mathbf{m}})',$$

$$\tilde{\mathbf{S}}_t = \sum_{i=1}^n (\tilde{\mathbf{x}}_i - \tilde{\mathbf{m}})(\tilde{\mathbf{x}}_i - \tilde{\mathbf{m}})',$$

gdzie $\tilde{\mathbf{m}}_k$ i $\tilde{\mathbf{m}}$ są odpowiednio średnią w k -tej klasie oraz średnią całkowitą w przestrzeni cech.

Znalezienie zmiennych dyskryminacyjnych w przestrzeni cech sprowadza się do rozwiązania następującego problemu optymalizacji

$$\mathbf{b}_{opt} = \arg \max \frac{\mathbf{b}' \tilde{\mathbf{S}}_b \mathbf{b}}{\mathbf{b}' \tilde{\mathbf{S}}_t \mathbf{b}}.$$

Jądrowe zmienne dyskryminacyjne

Wiemy, że istnieją współczynniki b_i takie, że $\mathbf{b}_{opt} = \sum_{i=1}^n b_i \tilde{\mathbf{x}}_i$. Stąd zagadnienie optymalizacji jest równoważne zagadnieniu

$$\mathbf{b}_{opt} = \arg \max \frac{\mathbf{b}' \mathbf{K} \mathbf{D} \mathbf{K} \mathbf{b}}{\mathbf{b}' \mathbf{K} \mathbf{K} \mathbf{b}}$$

i optymalne wektory \mathbf{b} są równe wektorom własnym odpowiadającym maksymalnym wartościom własnym w uogólnionym zagadnieniu własnym

$$\mathbf{K} \mathbf{D} \mathbf{K} \mathbf{b} = \lambda \mathbf{K} \mathbf{K} \mathbf{b},$$

$$\mathbf{K} = \mathbf{P} \tilde{\mathbf{K}} \mathbf{P},$$

gdzie $\tilde{\mathbf{K}} = (k_{ij})$ ($k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$) jest macierzą jądrową, $\mathbf{P} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n$ jest macierzą centrowania oraz macierz \mathbf{D} zdefiniowana jest następująco

$$D_{ij} = \begin{cases} \frac{1}{n_k}, & \text{jeżeli } \mathbf{x}_i \text{ oraz } \mathbf{x}_j \text{ należą do } k\text{-tej klasy;} \\ 0, & \text{poza tym.} \end{cases}$$

Niech $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_s]$ and $\mathbf{\Lambda} = [\lambda_1, \dots, \lambda_s]$. W zapisie macierzowym powyższy problem można zapisać

$$\mathbf{KDKB} = \mathbf{\Lambda KKB}.$$

Rozwiązanie uogólnionego zagadnienia własnego nastrocza pewne trudności, ponieważ obydwie macierze \mathbf{KDK} oraz \mathbf{KK} są nieujemnie określone. Istnieje szereg algorytmów pokonujących w różny sposób te trudności. Zaprezentujemy dwa z nich.

Algorytm 1

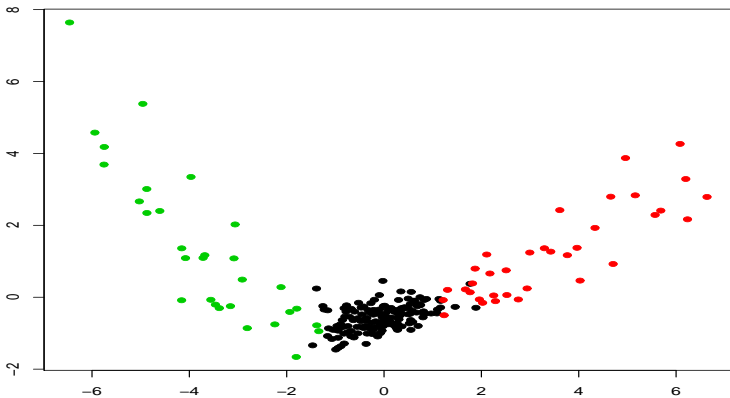
Najpopularniejszym ze sposobów uzyskania przybliżonego rozwiązania tego zagadnienia, wykorzystywanym również w regresji grzbietowej jest dokonanie regularyzacji macierzy $\mathbf{K}\mathbf{K}$, tzn. zastąpienie macierzy $\mathbf{K}\mathbf{K}$ nową nieosobliwą macierzą $\mathbf{K}\mathbf{K} + \epsilon\mathbf{I}$ (Friedman, 1989). Rozwiązujemy wówczas nowe uogólnione zagadnienie własne

$$(\mathbf{K}\mathbf{K} + \epsilon\mathbf{I})^{-1}\mathbf{K}\mathbf{D}\mathbf{K}\mathbf{B} = \lambda\mathbf{B},$$

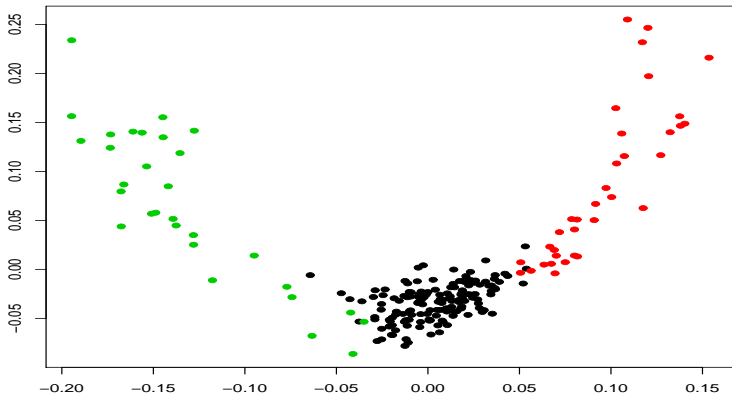
Rozwiązaniem tego zagadnienia są wartości i wektory własne macierzy

$$(\mathbf{K}\mathbf{K} + \epsilon\mathbf{I})^{-1}\mathbf{K}\mathbf{D}\mathbf{K}.$$

Rysunek: Rzut na przestrzeń 2 pierwszych klasycznych zmiennych dyskryminacyjnych



Rysunek: Rzut na przestrzeń 2 pierwszych jądrowych zmiennych dyskryminacyjnych - algorytm 1



Algorytm 2

Innym sposobem jest wykorzystanie odwrotności Moore'a-Penrosa $(\mathbf{K}\mathbf{K})^+$ macierzy $\mathbf{K}\mathbf{K}$. Przybliżone rozwiązanie jest wówczas postaci:

$$(\mathbf{K}\mathbf{K})^+ \mathbf{K}\mathbf{D}\mathbf{K}\mathbf{B} = \mathbf{\Lambda}\mathbf{B}.$$

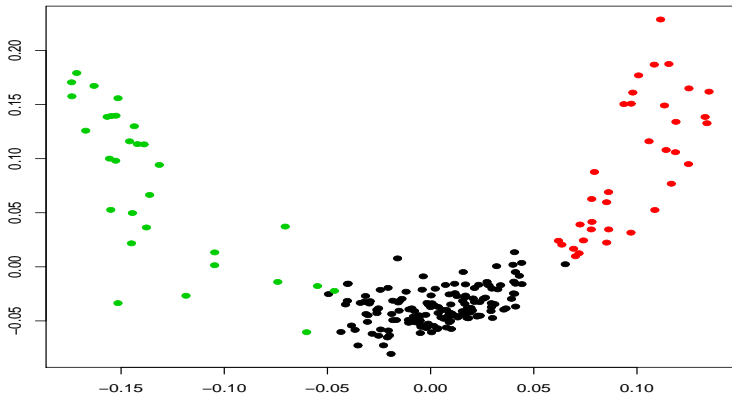
Wykorzystując fakt, że $(\mathbf{K}\mathbf{K})^+ \mathbf{K} = \mathbf{K}^+$ powyższe zagadnienie własne jest równoważne następującemu

$$\mathbf{K}^+ \mathbf{D}\mathbf{K}\mathbf{B} = \mathbf{\Lambda}\mathbf{B}.$$

Rozwiązaniem tego zagadnienia są wartości i wektory własne macierzy

$$\mathbf{K}^+ \mathbf{D}\mathbf{K}.$$





Rysunek: Rzut na przestrzeń 2 pierwszych jądrowych zmiennych dyskryminacyjnych - algorytm 2



Metoda	Błąd %	Czas s
Klasyczne	5,36	0,01
Algorytm 1	3,57	0,01
Algorytm 2	0,00	0,01

Metoda	Błąd %	Czas s
Klasyczne	24,44	0,03
Algorytm 1	21,54	13,77
Algorytm 2	17,52	20,64

- klasyczne zmienne dyskryminacyjne, **R**:
wynik=lda(Klasa~, data=ZbiorDanych)
predict(wynik)\$x
- jądrove zmienne dyskryminacyjne, **R**:
library(kernlab)
jądro gaussowskie: *rbfdot(sigma=c)*
jądro wielomianowe: *polydot(degree = 2, scale = a, offset = 1)*
- zagadnienie własne, **R**:
eigen(Macierz)
wartości własne: *eigen(Macierz)\$values*
wektory własne: *eigen(Macierz)\$vectors*

-  Baudat G. and Anouar F. (2000), Generalized discriminant analysis using a kernel approach, *Neural Computation* 12, 2385-2404.
-  Friedman J.H. (1989), Regularized discriminant analysis, *Journal of the American Statistical Association* 84, 165-175.
-  Krzyśko, M. (2009): *Podstawy wielowymiarowego wnioskowania statystycznego*, Wydawnictwo Naukowe UAM, Poznan.
-  Zhang Z., Dai G., Xu C. and Jordan M.I. (2010), Regularized discriminant analysis, ridge regression and beyond, *Journal of Machine Learning Research* 11, 2199-2228.

Funkcjonalne zmienne dyskryminacyjne

Tomasz Górecki, Mirosław Krzyśko, Łukasz Waszak

Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza

Wielowymiarowe metody statystyczne



Niech y_{lij} oznacza zaobserwowaną wartość badanej cechy statystycznej na i -tej jednostce pochodzącej z l -tej klasy w j -tym momencie czasowym, gdzie $i = 1, 2, \dots, N_l$, $j = 1, 2, \dots, J_i$, $l = 1, 2, \dots, L$, $N_1 + N_2 + \dots + N_L = N$. Momenty obserwacji t_{lij} 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 $\{t_{lij}, y_{lij}\}$, gdzie $t_{lij} \in I$, $i = 1, 2, \dots, N_l$, $j = 1, 2, \dots, J_i$, $l = 1, 2, \dots, L$.

Dane dyskretne $\{t_{lij}, y_{lij}\}$ możemy przekształcić na **dane funkcjonalne**:

$$\{x_{li}(t), i = 1, 2, \dots, N_l, l = 1, 2, \dots, L, t \in I\}.$$

Ponieważ proces przekształcania danych przeprowadzamy dla każdego $i = 1, 2, \dots, N_l$ oraz każdego $l = 1, 2, \dots, L$ oddzielnie, to dalsze rozważania dotyczą pojedynczej funkcji $x(t)$, $t \in I$.



Jednym ze sposobów wygładzania danych dyskretnych, do funkcji ciągłej $x(t)$ na przedziale I jest przedstawienie tej funkcji jako kombinacji liniowej N ortonormalnych funkcji bazowych φ_k :

$$x(t) = \sum_{k=0}^{N-1} c_k \varphi_k(t), \quad t \in I.$$

Współczynniki c_k tej kombinacji liniowej dobiera się metodą najmniejszych kwadratów, tj. tak aby minimalizowały funkcję:

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

gdzie $\mathbf{c} = (c_0, c_1, \dots, c_{N-1})'$ oraz Φ jest macierzą wymiaru $J \times N$ zawierającą wartości $\varphi_k(t_j)$. Stąd ocena wektora \mathbf{c} jest równa

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



Założmy, że obserwujemy realizację procesu stochastycznego $X(t) \in L_2(I)$, gdzie $L_2(I)$ jest przestrzenią funkcji całkowalnych z kwadratem na przedziale I , wyposażoną w iloczyn skalarny

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

Zakładamy również, że proces jest scentrowany oraz ma skończoną wariancję.



Zmienne dyskryminacyjne możemy zdefiniować zarówno dla wektorów losowych skończenie wymiarowych $X \in \mathbb{R}^p$, jak i dla procesów stochastycznych $X(t) \in L_2(I)$ w sposób następujący. Niech $H = \mathbb{R}^p$ w przypadku wektorowym oraz $H = L_2(I)$ w przypadku funkcjonalnym. Wówczas pierwsza wartość własna λ_1 i związany z nią wektor wag lub funkcja wagowa u_1 zdefiniowane są następująco:

$$\lambda_1 = \sup_{u \in H} \frac{u' S_B u}{u' S_W u} = \frac{u_1' S_B u_1}{u_1' S_W u_1}$$

oraz

$$u_1' S_W u_1 = 1,$$

gdzie $u' S_B u$ i $u' S_W u$ są odpowiednio wariancją między klasową i wariancją wewnątrz klasową iloczynu skalarne $\langle u, X \rangle$. Ograniczenie powyższe zapewnia, że wariancja pierwszej zmiennej dyskryminacyjnej $U_1 = \langle u_1, X \rangle$ jest jednostkowa.



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

$$\lambda_k = \sup_{u \in H} \frac{u' S_B u}{u' S_W u} = \frac{u_k' S_B u_k}{u_k' S_W u_k}$$

oraz

$$u_k' S_W u_l = \delta_{kl}.$$

Warunek ten zapewnia, że k -ta zmienna dyskryminacyjna $U_k = \langle u_k, X \rangle$ ma jednostkową wariancję i nie jest skorelowana z $k - 1$ pierwszymi zmiennymi dyskryminacyjnymi.



Można pokazać, że analiza zmiennych dyskryminacyjnych procesów $X(t)$ skończenie wymiarowych jest równoważna wielowymiarowej analizie zmiennych dyskryminacyjnych dla wektorów losowych \mathbf{c} . Analiza zmiennych dyskryminacyjnych dla wektorów losowych \mathbf{c} bazuje na macierzach $\mathbf{\Delta}$ i $\mathbf{\Gamma}$. W praktyce macierze te nie są znane. Oceniamy je na podstawie $N_1 + N_2 + \dots + N_L = N$ niezależnych realizacji wektora losowego \mathbf{c} zestawionych w L macierzy rozmiaru $N_l \times N$ postaci

$$\hat{\mathbf{C}}_l = \begin{bmatrix} \hat{c}_{l10} & \hat{c}_{l11} & \dots & \hat{c}_{l1(N-1)} \\ \hat{c}_{l20} & \hat{c}_{l21} & \dots & \hat{c}_{l2(N-1)} \\ \dots & \dots & \dots & \dots \\ \hat{c}_{lN_l 0} & \hat{c}_{lN_l 1} & \dots & \hat{c}_{lN_l(N-1)} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{c}}'_{l1} \\ \hat{\mathbf{c}}'_{l2} \\ \dots \\ \hat{\mathbf{c}}'_{lN_l} \end{bmatrix},$$

gdzie \hat{c}_{lrs} są ocenami uzyskanymi metodą najmniejszych kwadratów realizacji pochodzących z klasy l procesu losowego $X(t)$.



Z próby pochodzącej z l -tej klasy wyliczamy wektor średnich

$$\bar{\mathbf{c}}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \hat{\mathbf{c}}_{li}$$

oraz macierz

$$\hat{\mathbf{\Gamma}}_l = \frac{1}{N_l - 1} \sum_{i=1}^{N_l} (\hat{\mathbf{c}}_{li} - \bar{\mathbf{c}}_l)(\hat{\mathbf{c}}_{li} - \bar{\mathbf{c}}_l)', \quad l = 1, 2, \dots, L.$$

Z całej próby liczącej $N_1 + N_2 + \dots + N_L = N$ elementów wyliczamy oceny macierzy $\hat{\mathbf{\Delta}}$ i $\hat{\mathbf{\Gamma}}$:

$$\hat{\mathbf{\Delta}} = \frac{1}{L} \sum_{l=1}^L N_l \bar{\mathbf{c}}_l \bar{\mathbf{c}}_l',$$

$$\hat{\mathbf{\Gamma}} = \frac{1}{N - L} \sum_{l=1}^L (N_l - 1) \hat{\mathbf{\Gamma}}_l.$$



Następnie znajdujemy wartości własne $\hat{\lambda}_k$ i odpowiadające im wektory własne $\hat{\mathbf{u}}_k$ macierzy $\hat{\Gamma}^{-1}\hat{\Delta}$. Liczba niezerowych wartości własnych tej macierzy jest równa $\min(N, L - 1)$. Mając wyznaczone wektory własne $\hat{\mathbf{u}}_k$ odpowiadające niezerowym wartościom własnym wyznaczmy funkcje własne (funkcje wagowe):

$$\hat{u}_k(t) = \hat{\mathbf{u}}_k' \boldsymbol{\varphi}(t), \quad t \in I.$$



Stąd współrzędna rzutu i -tej realizacji $x_{li}(t)$ pochodzącej z l -tej klasy procesu $X(t)$ na j -tą funkcjonalną zmienną dyskryminacyjną jest równa

$$\hat{U}_{lij} = \langle \hat{u}_j(t), x_{li}(t) \rangle = \int \hat{u}_j(t) x_{li}(t) dt = \sum_{k=0}^{N-1} \hat{u}_{jk} \hat{c}_{lik} = \hat{\mathbf{c}}'_{li} \hat{\mathbf{u}}_j.$$



Przestrzeń \mathbb{R}^N wartości wektora losowego $\mathbf{c} = (c_0, c_1, \dots, c_{N-1})'$ przekształcamy za pomocą nieliniowej funkcji Ψ w przestrzeń Hilberta $H(k)$:

$$\Psi : \mathbb{R}^N \rightarrow H(k).$$

Niech $\Psi(\hat{c}_{lrs})$ będą realizacjami wektora \mathbf{c} przekształconymi za pomocą funkcji Ψ i scentrowanymi. Niech

$$\mathbf{m}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} \Psi(\hat{c}_{lj}),$$

$$\hat{\mathbf{W}}_l = \frac{1}{N_l - 1} \sum_{i=1}^{N_l} (\Psi(\hat{c}_{li}) - \mathbf{m}_l)(\Psi(\hat{c}_{li}) - \mathbf{m}_l)', \quad l = 1, 2, \dots, L,$$

$$\hat{\mathbf{B}} = \frac{1}{L} \sum_{l=1}^L N_l \mathbf{m}_l \mathbf{m}_l', \quad \hat{\mathbf{W}} = \frac{1}{N - L} \sum_{l=1}^L (N_l - 1) \hat{\mathbf{W}}_l.$$



Poszukujemy wektorów \mathbf{a}_i , które maksymalizują funkcję

$$J(\mathbf{a}_i) = \frac{\mathbf{a}_i' \hat{\mathbf{B}} \mathbf{a}_i}{\mathbf{a}_i' \hat{\mathbf{W}} \mathbf{a}_i},$$

przy dodatkowym warunku ograniczającym $\mathbf{a}_i' \hat{\mathbf{W}} \mathbf{a}_j = \delta_{ij}$, oznaczającym że zmienne dyskryminacyjne mają być nieskorelowane i mają mieć jednostkowe wariancje. Stosując trik jądrowy, stwierdzamy, że poszukujemy wektorów \mathbf{b}_i , które maksymalizują funkcję

$$J(\mathbf{b}_i) = \frac{\mathbf{b}_i' \mathbf{K} \mathbf{D} \mathbf{K} \mathbf{b}_i}{\mathbf{b}_i' \mathbf{K} (\mathbf{I} - \mathbf{D}) \mathbf{K} \mathbf{b}_i},$$

przy warunku ograniczającym $\mathbf{b}_i' \mathbf{K} (\mathbf{I} - \mathbf{D}) \mathbf{K} \mathbf{b}_j = \delta_{ij}$, gdzie \mathbf{K} jest macierzą jądrową, natomiast $\mathbf{D} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_L)$ jest macierzą blokową, w której blok \mathbf{D}_l jest macierzą stopnia N_l o elementach $\frac{1}{N_l}$, $l = 1, 2, \dots, L$.



Wydajność opisanych metod została przetestowana na danych Olive011 pochodzących z repozytorium UCR. Zbiór ten składa się z 60 szeregów czasowych o długości 570 (widma podczerwieni) podzielonych pomiędzy 4 klasy (kraje pochodzenia oliwy: Grecja (10), Włochy (17), Portugalia (8) i Hiszpania (25)). Zbiór testowy oraz uczący liczą po 30 obserwacji. Dyskretne szeregi czasowe zostały scentrowane, a następnie przekształcone w funkcje ciągłe. Użyta została baza ortonormalna FOURIERA:

$$\begin{aligned}\varphi_0(x) &= \frac{1}{\sqrt{T}}, \\ \varphi_{2k-1}(x) &= \frac{\sqrt{2}}{\sqrt{T}} \sin \frac{2\pi kx}{T}, \quad k = 1, 2, \dots \\ \varphi_{2k}(x) &= \frac{\sqrt{2}}{\sqrt{T}} \cos \frac{2\pi kx}{T}, \quad k = 1, 2, \dots\end{aligned}$$

Użyte zostało jądro wielomianowe postaci $f(\mathbf{x}, \mathbf{y}) = (1 + \mathbf{x}'\mathbf{y})^2$.



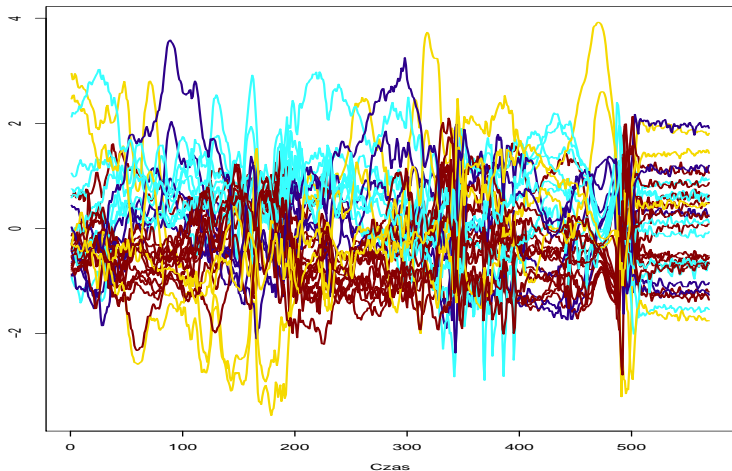
Najczęściej badane obiekty są przedstawiane jako punkty na wykresie w układzie dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych. W takim przypadku **kryterium jakości** skonstruowanych funkcjonalnych zmiennych dyskryminacyjnych ma postać:

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

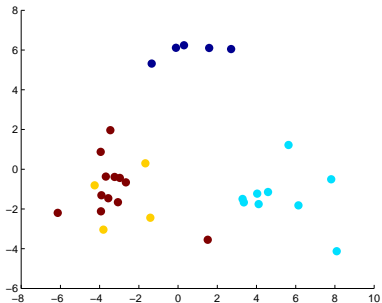
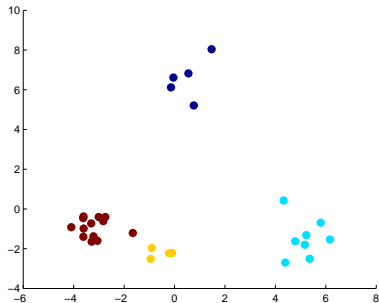
gdzie $\lambda_1 \geq \lambda_2 \geq \dots$ są niezerowymi wartościami własnymi. Im większa wartość wyrażenia tym więcej zmienności zachowują dwie pierwsze funkcjonalne zmienne dyskryminacyjne.



Rysunek : Zbiór uczący



Rysunek : Rzut danych na przestrzeń dwóch pierwszych funkcjonalnych zmiennych dyskryminacyjnych, po lewej — zbiór uczący, po prawej — testowy



Rysunek : Rzut danych na przestrzeń dwóch pierwszych jądrowych funkcjonalnych zmiennych dyskryminacyjnych, po lewej — zbiór uczący, po prawej — testowy

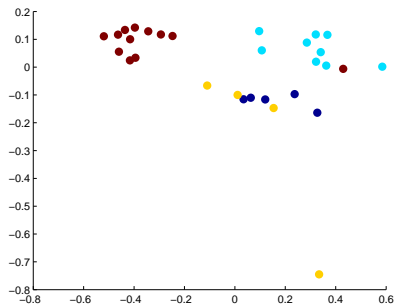
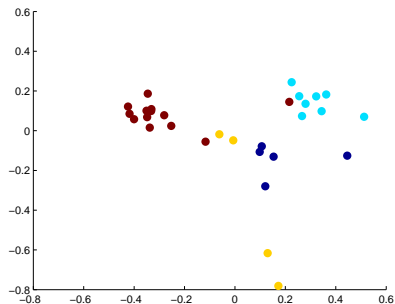


Tabela : Wariancja wyjaśniana przez dwie pierwsze funkcjonalne zmienne dyskryminacyjne

FZD	KFZD
73,11%	78,07%

