AUTOREFERAT

1. Imię i Nazwisko: Waldemar Wołyński

2. Specjalność naukowa: statystyka matematyczna

3. Dyplomy i stopnie naukowe:

Magister matematyki: 1986;
Wydział Matematyki i Fizyki Uniwersytetu im. Adama Mickiewicza;
Tytuł pracy magisterskiej: Jednostajne i lipschitzowskie homeomorfizmy przestrzeni Banacha;
Promotor: prof. dr hab. Lech Drewnowski.
Doktor: 1993;
Instytut Matematyki Uniwersytetu im. Adama Mickiewicza; stopień doktora nauk matematycznych w zakresie matematyki;
Tytuł rozprawy doktorskiej: Zagadnienia estymacji w regułach klasyfikacji grupowej;
Promotor: prof. dr hab. Mirosław Krzyśko.

4. Informacje o dotychczasowym zatrudnieniu:

1986 do dnia dzisiejszego: Uniwersytet im. Adama Mickiewicza w Poznaniu; Wydział Matematyki i Informatyki; Zakład Rachunku Prawdopodobieństwa i Statystyki Matematycznej.
2000 - 2004: Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu; Wydział Zarządzania.

5. Osiągnięcie naukowe, o którym mowa w art. 16 ust. 2 ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. Nr 65, poz. 595, z późn. zm.): Jednotematyczny cykl 8 prac pt.

Odległości probabilistyczne w analizie dyskryminacyjnej.

6. Omówienie celu naukowego i osiągniętych wyników:

1 Wstęp

Przyglądając się klasycznemu, pochodzącemu od R.A. Fishera (1936), rozwiązaniu problemu analizy dyskryminacyjnej widać, że używany przez niego klasyfikator (LDA) przyporządkowuje obiekt do populacji najbliższej mu w sensie odległości Mahalanobisa. Jeżeli przyjmiemy dodatkowe założenie o normalności rozkładów populacji z jednakowymi macierzami kowariancji, to również prawdopodobieństwa błędnych klasyfikacji wyrażają się poprzez odległość Mahalanobisa.
Model zakładający jednorodność macierzy kowariancji i/lub normalność rozkładów populacji jest mocno restrykcyjny. Wydaje się zatem naturalne rozważanie takich procedur klasyfikacyjnych, które nie mając tych krępujących założeń nadal zachowują ideę klasyfikacji obiektu do populacji najbliższej mu w sensie zadanej odległości.
W literaturze znaleźć można wiele różnych pomysłów na wprowadzenie odległości w zbiorze miar probabilistycznych. W swoich pracach używałem (najczęściej rozważanych) odległości: Chernoffa, Morisity oraz Kullbacka-Leiblera.
Niech zatem $(\Omega, \mathcal{F}, \mu)$ będzie przestrzenią z miarą, $\mathcal{P}$ zbiorem rozkładów prawdopodobieństwa na $\mathcal{F}$ absolutnie ciągłych ze względu na miarę μ.
Ponadto, niech f_1, f_2 będą gęstościami rozkładów prawdopodobieństw P_1, P_2 ze zbioru $\mathcal{P}$.
Zdefiniujmy następujące odległości probabilistyczne na zbiorze Ω.

- Odległość Chernoffa (1952):

$$\rho_C(f_1, f_2) = -\ln \int_\Omega f_1^{1-s} f_2^s d\mu, \;\; s \in [0,1].$$

  Szczególny przypadek tej odległości dla $s = \frac{1}{2}$ nazywany jest odległością Bhattacharyya (1943) - oznaczenie ρ_B.

- Odległość Morisity (1959):

$$\rho_M(f_1, f_2) = -\ln \frac{2\Delta(f_1, f_2)}{\Delta(f_1, f_1) + \Delta(f_2, f_2)},$$

  gdzie

$$\Delta(f_1, f_2) = \int_\Omega f_1 f_2 d\mu.$$

  Odległość ta jest ściśle związana z klasyczną odległością w przestrzeni $L^2(\Omega)$.

- Odległość Kullbacka-Leiblera (1951):

$$\rho_{KL}(f_1, f_2) = \int_\Omega [f_1 - f_2] \ln \frac{f_1}{f_2} d\mu.$$

2 Osiągnięcie naukowe

Osiągnięcie naukowe stanowi 8 następujących prac:

H1. M. Krzyśko, W. Wołyński, *Bayes rules and minimum distance rules in the statistical group classification problems*, Discussiones Mathematicae. Algebra and Stochastic Methods **15** (1995), 313-323.

H2. M. Krzyśko, W. Wołyński, *Discriminant rules based on distances*, Tatra Mountains Math. Publ. **7** (1996), 289-296.

H3. M. Krzyśko, W. Wołyński, *Linear Discriminant Functions for Stationary Time Series*, Biometrical Journal **39** (1997), 955-973. IF: 1.614.

H4. M. Krzyśko, W. Wołyński, *Pairwise multiple discriminant analysis*, Advances and Applications in Statistics **2** (2002), 249-268.

H5. W. Wołyński, *Minimal sample size in the group classification problem*, Journal of Classification, **22** (2005), 49-58. IF: 1.413.

H6. R.Kala, M. Krzyśko, W. Wołyński, *Two preliminary tests for discriminant analysis*, Communications in Statistics - Simulation and Computation **34** (2005), 179-189. IF: 0,482.

H7. M. Krzyśko, W. Wołyński, *New variants of pairwise classification*, European Journal of Operational Research **199** (2009), 512-519. IF: 2.277.

H8. W. Wołyński, *Kernel linear discriminant functions*, In: Data analysis methods and its applications, J. Pociecha, R. Decker (Ed.), C.H. Beck, Warszawa 2012, 59-70.

Omawiane są w nich zastosowania odległości probabilistycznych w różnych aspektach analizy dyskryminacyjnej. W pracy [H1] zastosowano metodę minimalizacji odległości probabilistycznych do konstrukcji klasyfikatorów w zagadnieniu klasyfikacji grup obiektów. Praca [H2] poświęcona jest rozwiązaniu problemu wyznaczania optymalnych liniowych procedur klasyfikacyjnych maksymalizujących odległości pomiędzy populacjami. W pracy [H3] zastosowano wyniki uzyskane w pracy [H2] do zagadnień klasyfikacji stacjonarnych

szeregów czasowych. Podano w niej również twierdzenia mówiące o asymptotycznych własnościach uzyskanych rozwiązań. W pracach [H2] i [H3] rozważany był jedynie przypadek dwóch populacji. W następnej pracy [H4] podano sposób wykorzystania tych metod do zagadnień z dowolną liczbą populacji. W pracy [H7] zebrano i porównano wiele procedur łączenia klasyfikatorów binarnych. Procedury te mają szczególne znaczenie, gdyż klasyfikatory wykorzystujące odległości probabilistyczne pomiędzy populacjami są "z natury" binarne. W pracy [H7] zaproponowano również szereg modyfikacji istniejących procedur. Praca [H5] podaje oszacowanie minimalnej liczebności klasyfikowanej grupy obiektów zapewniającej ustalony poziom ryzyka bayesowskiego. Oszacowanie to podane jest w terminach odległości Chernoffa. W pracy [H6] wchodzącej w skład mojej rozprawy habilitacyjnej rozważane jest zagadnienie testów wstępnych w analizie dyskryminacyjnej. Za pomocą tych testów weryfikujemy hipotezę mówiącą, że analizowany przez nas nowy obiekt (kandydat) nie należy do żadnej populacji z rozważanej klasy (standardy). W pracy [H8] podano postaci tzw. jądrowych klasyfikatorów opartych na maksymalizacji odległości probabilistycznych przeniesionych z przestrzeni $\mathbf{R}^p$ do przestrzeni Hilberta z jądrem reprodukującym. Optymalnym liniowym klasyfikatorom w przestrzeni Hilberta odpowiadają nieliniowe powierzchnie rozdzielające populacje w pierwotnej przestrzeni cech. Zastosowana w tym celu technika zwana "kernel trick" jest aktualnie szeroko stosowana w wielu algorytmach statystycznych.

2.1 Klasyfikacja grup obiektów

Klasyczne zagadnienie analizy dyskryminacyjnej polega na konstrukcji algorytmu pozwalającego na klasyfikację obiektu opisanego wektorem $\boldsymbol{x}_0$ obserwowanych cech, do jednej z k populacji (klas). W pracy z 1980 roku Abusev i Lumelsky rozważali problem klasyfikacji nie pojedynczego obiektu lecz grupy obiektów pochodzących z tej samej populacji i opisanych za pomocą N_0 niezależnych wektorów $x_{01}, x_{02}, \ldots, x_{0N_0}$.
Problemem tym zajmowałem się w szeregu prac opublikowanych w latach 1992-1996. W pracy [H1] rozważałem zagadnienie klasyfikacji grupy $\boldsymbol{x}_0 = (x_{01}, x_{02}, \ldots, x_{0N_0})$ niezależnych obserwacji do jednej z k populacji o p–wymiarowych rozkładach normalnych $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ $(i = 1, 2, \ldots k)$, w przestrzeni statystyk dostatecznych

$$\bar{x}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}, \quad A_i = \sum_{j=1}^{N_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)', \quad i = 0, 1, 2, \ldots, k.$$

Reguła klasyfikacyjna oparta na odległościach probabilistycznych polega na

klasyfikawaniu $\boldsymbol{x}_0$ do populacji $N_p(\boldsymbol{\mu}_{i_0}, \boldsymbol{\Sigma}_{i_0})$ wtedy i tylko wtedy, gdy

$$\rho_j(f_0, f_{i_0}) = \min_{1 \leqslant i \leqslant k} \rho_j(f_0, f_i),$$

dla $j \in \{B, M, KL\}$.
W pracy [H1] wyprowadzono jawne postaci występujących w regule odległości. Podaje je poniższe twierdzenie.

Twierdzenie 1. *Niech f_1 i f_2 oznaczają funkcje gęstości statystyk dostatecznych z populacji o rozkładach $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ odpowiednio. Wtedy*

$$\begin{aligned}\rho_B(f_1, f_2) = &-\frac{p}{4}\ln(N_1N_2) - \frac{p(N_1+N_2)}{4}\ln 2 - \ln\Gamma_p\Big(\frac{N_1+N_2-2}{4}\Big)\\ &+ \frac{1}{2}\ln\Big[\Gamma_p\Big(\frac{N_1-1}{2}\Big)\Gamma_p\Big(\frac{N_2-1}{2}\Big)\Big] - \frac{N_2}{4}\ln|\boldsymbol{\Sigma}_1| - \frac{N_1}{4}\ln|\boldsymbol{\Sigma}_2|\\ &+ \frac{N_1+N_2-2}{4}\ln|\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2| + \frac{1}{2}\ln|N_1\boldsymbol{\Sigma}_2+N_2\boldsymbol{\Sigma}_1|\\ &+ \frac{N_1N_2}{4}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(N_1\boldsymbol{\Sigma}_2+N_2\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),\end{aligned}$$

$$\rho_M(f_1, f_2) = -\ln\frac{2\Delta(f_1, f_2)}{\Delta(f_1, f_1)+\Delta(f_2, f_2)},$$

gdzie

$$\begin{aligned}\Delta(f_1, f_2) = &\pi^{-\frac{p}{2}}(N_1N_2)^{\frac{p}{2}}2^{-\frac{p(p+2)}{2}}\Gamma_p\Big(\frac{N_1+N_2-p-3}{2}\Big)\\ &\times\Gamma_p^{-1}\Big(\frac{N_1-1}{2}\Big)\Gamma_p^{-1}\Big(\frac{N_2-1}{2}\Big)\\ &\times|\boldsymbol{\Sigma}_1|^{\frac{N_2-p-2}{2}}|\boldsymbol{\Sigma}_2|^{\frac{N_1-p-2}{2}}|N_1\boldsymbol{\Sigma}_2+N_2\boldsymbol{\Sigma}_1|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2|^{-\frac{N_1+N_2-p-3}{2}}\\ &\times\exp\Big[-\frac{N_1N_2}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(N_1\boldsymbol{\Sigma}_2+N_2\boldsymbol{\Sigma}_1)^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\Big],\end{aligned}$$

oraz

$$\begin{aligned}&\rho_{KL}(f_1, f_2)\\ &= \frac{1}{2}\Big[(N_1-N_2)\sum_{j=1}^{p}\Big(\psi\Big(\frac{N_1-j}{2}\Big)-\psi\Big(\frac{N_2-j}{2}\Big)\Big) + (N_1-N_2)\ln(|\boldsymbol{\Sigma}_1|/|\boldsymbol{\Sigma}_2|)\\ &+ N_2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) + N_1(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\ &+ \Big(\frac{N_2}{N_1}+N_1-1\Big)\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}) + \Big(\frac{N_1}{N_2}+N_2-1\Big)\mathrm{tr}(\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}) - (N_1+N_2)p\Big].\end{aligned}$$

W pracy tej, za pomocą symulacji porównano klasyczną kwadratową regułę klasyfikacyjną z regułami minimalizującymi odległości probabilistyczne. Z symulacji tych wynika, że reguły oparte na minimalizacji odległości probabilistcznych są zazwyczaj lepsze od klasycznych reguł bayesowskich.

2.2 Liniowe reguły klasyfikacyjne maksymalizujące odległości probabilistyczne pomiędzy populacjami

Szczególne miejsce wśród procedur klasyfikacyjnych zajmują procedury liniowe, tzn. takie w których powierzchnia rozdzielająca populacje jest hiperpłaszczyzną. Własność tę posiada optymalna - bayesowska reguła klasyfikacyjna w modelu normalnym z jednakowymi macierzami kowariancyjnymi. Opuszczenie założenia jednorodności macierzy kowariancyjnych sprawia, że w rozwiązaniu bayesowskim powierzchnia rozdzielająca jest hiperpowierzchnią stopnia drugiego. Poszukiwanie innych niż bayesowskie lecz liniowych procedur w przypadku różnych macierzy kowariancji zapoczątkowała praca Andersona i Bahadura (1962). Pierwszą pracą wykorzystującą odległość Bhattacharyya do konstrukcji procedur liniowych była praca Chaudhuri, Borwankar i Rao (1991a). Zagadnienie procedur liniowych dla stacjonarnych szeregów czasowych badane było przez Schumwaya i Ungera (1974) oraz Chaudhuri, Borwankar i Rao (1991b). Tematykę tę rozwijałem w pracach opublikowanych w latach 1996-2002, w szczególności elementem mojego osiągnięcia naukowego są poświęcone temu tematowi prace [H2], [H3] i [H4].
Praca [H2] zawiera twierdzenia podające procedury wyznaczania wektora $\boldsymbol{a}$ oraz stałej c hiperpłaszczyzny $\boldsymbol{a'x} - c = 0$ rozdzielającej populacje w p-wymiarowym modelu normalnym, maksymalizujące odległości Chernoffa, Morosity i Kulbacka-Leiblera pomiędzy tymi populacjami.

Twierdzenie 2. *Niech dane będą dwie populacje o rozkładach* $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ *oraz niech* $\boldsymbol{\Sigma}_1 = \boldsymbol{P'P}$, $\boldsymbol{\Sigma}_2 = \boldsymbol{P'\Lambda P}$, *gdzie* $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ *oraz* λ_i $(1 = 1, \ldots, p)$ *są wartościami własnymi macierzy* $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}$.
Wtedy wektor $\boldsymbol{a}$ *ma postać*

$$\boldsymbol{a} = (\boldsymbol{\Sigma}_1 + \theta\boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

gdzie θ *jest punktem stałym odwzorowania* ψ_i $(i = C, M, KL)$ *postaci*

$$\psi_C(\theta) = \frac{s}{1-s}\frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}} + \Big(\frac{s}{1-s} - \theta\Big)A_1(\boldsymbol{\beta}'\boldsymbol{\beta}),$$
$$A_1 = \frac{s(1-s)(\boldsymbol{\beta}'\boldsymbol{\eta})^2}{[(1-s)\boldsymbol{\beta}'\boldsymbol{\beta} + s\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}]^2} - \frac{1}{(1-s)\boldsymbol{\beta}'\boldsymbol{\beta} + s\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}},$$
$$\psi_M(\theta) = \Big[\frac{B_2}{A_2}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-\frac{1}{2}} - (A_2\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}\Big]\theta + 1 - \frac{B_2}{A_2}(\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})^{-\frac{1}{2}} + [A_2(\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})]^{-1},$$
$$A_2 = \frac{(\boldsymbol{\beta}'\boldsymbol{\eta})^2}{[\boldsymbol{\beta}'(I+\boldsymbol{\Lambda})\boldsymbol{\beta}]^2} - (\boldsymbol{\beta}'(\boldsymbol{I}+\boldsymbol{\Lambda})\boldsymbol{\beta})^{-1},\ \ B_2 = \Big((\boldsymbol{\beta}'\boldsymbol{\beta})^{\frac{1}{2}} + (\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})^{\frac{1}{2}}\Big)^{-1},$$
$$\psi_{KL}(\theta) = A_3(\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})\theta - B_3(\boldsymbol{\beta}'\Lambda\boldsymbol{\beta}) + (\boldsymbol{\beta}'\Lambda\boldsymbol{\beta})(\boldsymbol{\beta}'\beta)^{-1},$$
$$A_3 = \frac{(\boldsymbol{\beta}'\boldsymbol{\eta})^2 + \boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}}{(\boldsymbol{\beta}'\boldsymbol{\beta})^2},\ \ B_3 = \frac{(\boldsymbol{\beta}'\boldsymbol{\eta})^2 + \boldsymbol{\beta}'\boldsymbol{\beta}}{(\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})^2},$$

przy czym $\boldsymbol{\beta} = \boldsymbol{P}\boldsymbol{a}$ *oraz* $\boldsymbol{\eta} = (\boldsymbol{I} - \theta\boldsymbol{\Lambda})\boldsymbol{\beta}$.

W pracy [H2] pokazano również, że jeżeli stała θ z twierdzenia 2 jest taka, że macierz $\boldsymbol{\Sigma}_1 + \theta\boldsymbol{\Sigma}_2$ jest dodatnio określona, a stała c ma postać:

$$c = \boldsymbol{a}'\boldsymbol{\mu}_1 - \boldsymbol{a}'\boldsymbol{\Sigma}_1\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{\mu}_2 + \theta(\boldsymbol{a}'\boldsymbol{\Sigma}_2\boldsymbol{a}),$$

to otrzymana w ten sposób liniowa procedura klasyfikacyjna jest dopuszczalna w klasie procedur liniowych.

Praca [H3] rozszerza wyniki z pracy [H2] na problem dyskryminacji w zagadnieniach stacjonarnych szeregów czasowych. Podano w niej twierdzenia mówiące o granicznych własnościach procedur liniowych opartych o odległości probabilistyczne.

Przy założeniu warunków regularności podano dowody następujących twierdzeń:

Twierdzenie 3. *Niech g_1 i g_2 oznaczają funkcje gęstości funkcji dyskryminacyjnej $y = \boldsymbol{a}'\boldsymbol{x}$ w populacjach określonych rozkładami $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ odpowiednio oraz niech h_1 i h_2 oznaczają gęstości spektralne.*
Jeżeli $h_\theta(\lambda) = h_1(\lambda) + \theta h_2(\lambda) > 0$ dla $\lambda \in [-\pi, \pi]$, to

$$\lim_{T\to\infty} T^{-1}\rho_C(g_1, g_2) = \frac{1}{4}G(\theta),\ \ s = \frac{1}{2},$$
$$\lim_{T\to\infty} T^{-1}\rho_M(g_1, g_2) = \frac{1}{2}G(\theta),$$
$$\lim_{T\to\infty} T^{-1}\rho_{KL}(g_1, g_2) = \frac{1}{2}H(\theta),$$

gdzie

$$G(\theta) = \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{dM(\lambda)}{h_\theta(\lambda)}\Big)^2 / \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_1(\lambda)+h_2(\lambda)}{h^2_{\theta_i}(\lambda)}dM(\lambda)\Big),$$

$$H(\theta) = \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{dM(\lambda)}{h_\theta(\lambda)}\Big)^2\Big[\Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_1(\lambda)}{h^2_\theta(\lambda)}dM(\lambda)\Big)^{-1} + \Big(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{h_2(\lambda)}{h^2_{\theta_3}(\lambda)}dM(\lambda)\Big)^{-1}\Big].$$

Twierdzenie 4. *Funkcje $G(\theta)$ i $H(\theta)$ z twierdzenia 3 mają maksima globalne w punkcie $\theta = 1$.*

Zatem pokazano, że asymptotycznie optymalna hiperpłaszczyzna rozdzielająca ma postać $\boldsymbol{a}'_\infty \boldsymbol{x} - c_\infty = 0$, gdzie

$$\boldsymbol{a}_\infty = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

oraz

$$c_\infty = \boldsymbol{a}_\infty{}'\boldsymbol{\mu}_1 - \boldsymbol{a}_\infty{}'\boldsymbol{\Sigma}_1\boldsymbol{a}_\infty = \boldsymbol{a}_\infty{}'\boldsymbol{\mu}_2 + \boldsymbol{a}_\infty{}'\boldsymbol{\Sigma}_2\boldsymbol{a}_\infty.$$

2.3 Metody łączenia klasyfikatorów binarnych

Omawiane w pracach [H2] i [H3] liniowe procedury dotyczyły jedynie przypadku dwóch populacji (klas). Uzyskane z takich procedur klasyfikatory nazywamy klasyfikatorami binarnymi. Niech teraz liczba klas $k > 2$. Zagadnienia klasyfikacyjne z liczbą klas większą od dwóch nazywamy wieloklasowymi. Interesujące jest również zagadnienie odwrotne polegające na dekompozycji zagadnienia wieloklasowego na zagadnienia binarne. Nawet gdy istnieje "naturalne" rozwiązanie wieloklasowe, to zazwyczaj przypadek dwóch klas jest najprostszy obliczeniowo. W tej sytuacji sensowne wydaje się "sprowadzenie" jednego dużego zagadnienia wieloklasowego do być może wielu, ale jednak o wiele prostszych, zagadnień binarnych. Jednymi z pierwszych prac dotyczących tej tematyki były prace Hastiego i Tibshiraniego (1996, 1998). W pracy [H4] podano algorytm pozwalający na wykorzystanie binarnych klasyfikatorów liniowych maksymalizujących odległości pomiędzy klasami w zagadnieniach wieloklasowych. Algorytm ten składa się z dwóch kroków. W pierwszym kroku, dla każdej z $k(k-1)/2$ par różnych populacji wyznaczamy estymator liniowej funkcji klasyfikującej u_{ij} postaci:

$$u_{ij}(\boldsymbol{x}) = (\hat{\boldsymbol{a}}'_{ij}\boldsymbol{x} - \hat{c}_{ij})/|\hat{\boldsymbol{a}}_{ij}|,$$

gdzie

$$\begin{aligned}\hat{\boldsymbol{a}}_{ij} &= (\boldsymbol{S}_i + \theta \boldsymbol{S}_j)^{-1}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_j),\\ \hat{c}_{ij} &= \hat{\boldsymbol{a}}'_{ij}\bar{\boldsymbol{x}}_i - \hat{\boldsymbol{a}}'_{ij}\boldsymbol{S}_i\hat{\boldsymbol{a}}_{ij} = \hat{\boldsymbol{a}}'_{ij}\bar{\boldsymbol{x}}_j + \theta(\hat{\boldsymbol{a}}'_{ij}\boldsymbol{S}_j\hat{\boldsymbol{a}}_{ij}),\\ |\hat{\boldsymbol{a}}_{ij}| &= \sqrt{\hat{\boldsymbol{a}}'_{ij}\hat{\boldsymbol{a}}_{ij}},\ \ i,j = 1,2,\ldots,k,\ i \neq j.\end{aligned}$$

Następnie klasyfikowanej obserwacji $\boldsymbol{x}_0$ przyporządkowujemy prawdopodobieństwa a posteriori wykorzystując tzw. funkcję sigmoidalną postaci:

$$\hat{p}_{ij}(\boldsymbol{x}_0) = 1/[1 + \exp\{u_{ij}(\boldsymbol{x}_0)\}].$$

W kroku drugim szacujemy prawdopodobieństwa a posteriori $p_i(\boldsymbol{x}_0)$ przynależności obiektu do każdej z rozważanych populacji wykorzystując algorytm Hastiego i Tibshiraniego.
Mamy

$$p_{ij}(\boldsymbol{x}_0) = \frac{p_i(\boldsymbol{x}_0)}{p_i(\boldsymbol{x}_0) + p_j(\boldsymbol{x}_0)},\ i < j, \quad (p_{ji}(\boldsymbol{x}_0) = 1 - p_{ij}(\boldsymbol{x}_0)).$$

Bezpośrednie wykorzystanie powyższej zależności do oszacowania prawdopodobieństw $p_i(\boldsymbol{x}_0)$ jest, niestety, niemożliwe. Prowadzi ono bowiem do rozwiązania, zazwyczaj sprzecznego, układu $k(k-1)/2$ równań z k niewiadomymi. Prawdopodobieństwa a posteriori $p_i(\boldsymbol{x}_0)$ szacujemy poprzez minimalizację odległości Kullbacka-Leiblera postaci

$$\rho_{KL}(p_1(\boldsymbol{x}_0),\ldots,p_k(\boldsymbol{x}_0)) = \sum_{j\neq i} n_{ij}\left[\hat{p}_{ij}(\boldsymbol{x}_0)\log\frac{\hat{p}_{ij}(\boldsymbol{x}_0)}{p_{ij}(\boldsymbol{x}_0)}\right],$$

gdzie liczebności $n_{ij} = n_i + n_j$ pełnią funkcję wag. Do rozwiązania zagadnienia minimalizacyjnego stosujemy algorytm Bradleya-Terry'ego.
Uzyskane w ten sposób oszacowania prawdopodobieństw a posteriori są (w sensie klasyfikacji) równoważne z dużo prostszymi oszacowaniami postaci:

$$\tilde{p}_i(\boldsymbol{x}_0) = \frac{2}{k(k-1)}\sum_{j\neq i}\hat{p}_{ij}(\boldsymbol{x}_0).$$

Następnie, w klasyczny sposób, obiekt klasyfikujemy do populacji o największym oszacowanym prawdopodobieństwie a posteriori.
Procedura Hastiego i Tibshiraniego ma jedną zasadniczą wadę. Klasyfikując obserwację $\boldsymbol{x}_0$ musimy, za pomocą klasyfikatorów binarnych, oszacować $k(k-1)/2$ prawdopodobieństw $p_{ij}(\boldsymbol{x}_0)$. Zauważmy jednak, że jeżeli klasyfikowana

obserwacja w rzeczywistości pochodzi z i-tej klasy, to w oszacowaniu tych prawdopodobieństw obserwacje uczące z i-tej klasy biorą udział tylko $k-1$ razy. Do pozostałych $(k-1)(k-2)/2$ oszacowań, obserwacje uczące z i-tej klasy nie są w ogóle wykorzystywane, co oznacza, że ich wynik może być w zasadzie zupełnie dowolny.
W pracy [H7] zaproponowano szereg modyfikacji procedury Hastiego i Tibshiraniego oraz innych procedur łączenia klasyfikatorów binarnych. W szczególności rozważano procedury pochodzące z prac Moreiry i Mayoraza (1998) oraz Jelonka i Stefanowskiego (1998). Główna idea poprawy algorytmów polega na wprowadzeniu w procesie estymacji prawdopodobieństw $p_i(\boldsymbol{x}_0)$ wag uzależnionych od tego, czy obserwacja $\boldsymbol{x}_0$ należy do klasy i-tej lub j-tej, czy raczej do jednej z pozostałych klas.
Praca [H7] zawiera również wyniki obszernych badań symulacyjnych porównujących jakość 16 procedur łączenia klasyfikatorów. Do badania istotności różnic użyto testu Imana i Davenporta (1980), a do połączenia procedur w grupy jednorodne zastosowano procedurę Nemenyiego (1963).

2.4 Wyznaczanie minimalnej liczebności klasyfikowanej grupy

W październiku 1999 roku nawiązałem współpracę naukową z pracownikami Katedry Geotechniki Uniwersytetu Przyrodniczego w Poznaniu kierowanej przez prof. dra hab. Zbigniewa Młynarka. W opracowywanych przez pracowników Katedry Geotechniki metodach analizy gruntów w oparciu o technikę sondowania statycznego znalazło zastosowanie szereg procedur statystyki wielowymiarowej, a w szczególności najbardziej mnie interesujące procedury analizy dyskryminacyjnej. Wyniki wspólnych badań zostały opublikowane w szeregu prac z lat 2001-2008. Wśród nich znajduje się praca [H5] podająca rozwiązanie postawionego mi przez prof. Młynarka problemu wyznaczania minimalnej liczby sondowań potrzebnych do prawidłowej klasyfikacji gruntu. Zagadnienie to w języku analizy dyskryminacyjnej sprowadza się do wyznaczenia minimalnej liczby N_0, przy której klasyfikowana grupa $\boldsymbol{x}_0 = (x_{01}, x_{02}, \ldots, x_{0N_0})$ niezależnych obserwacji spełnia zadane kryterium optymalności.
Rozważmy przypadek dwóch populacji opisanych gęstościami f_1, f_2 odpowiednio. Ponadto przyjmujemy, że q_1, q_2 są prawdopodobieństwami a priori przynależności grupy $\boldsymbol{x}_0$ do tych populacji oraz jako kryterium optymalności przyjmujemy ryzyko bayesowskie. Założenia te prowadzą do wyboru minimalnej liczebności grupy N_0 w taki sposób, aby spełniona była nierówność

$$(*) \qquad \int_{\Omega} \min(q_1 f_1, q_2 f_2) d\mu \leqslant \alpha, \quad (0 < \alpha < 0.5),$$

gdzie α jest zadanym poziomem ryzyka.
W pracy [H5] podano procedury pozwalające na dokładne wyznaczenie optymalnej wielkości grupy w przypadku gdy populacje mają p-wymiarowe rozkłady normalne.
Podano również oszacowanie górne ryzyka bayesowskiego w dowolnym modelu z wykorzystaniem odległości Chernoffa. Pozwala ono na oszacowanie minimalnej liczebności grupy. Uzyskany wynik zawiera poniższe twierdzenie

Twierdzenie 5. *Dla minimalnej liczebności grupy N_0 spełniającej warunek (*), zachodzi nierówność*

$$N_0 \geqslant \frac{\ln\left(\frac{q_1^s q_2^{1-s}}{\alpha}\right)}{\rho_C(f_1, f_2)},\ 0 \leqslant s \leqslant 1.$$

Przykładowo dla dwóch populacji o p wymiarowych rozkładach normalnych $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ odpowiednio, minimalne N_0 wyznaczamy ze wzoru

$$N_0 = \min_{0 \leqslant s \leqslant 1} \lceil \ln\left(\frac{q_1^s q_2^{1-s}}{\alpha}\right) / \rho_C(s) \rceil,$$

gdzie

$$\rho_C(s) = \frac{s(1-s)}{2}(\mu_1 - \mu_2)'[s\boldsymbol{\Sigma}_1 + (1-s)\boldsymbol{\Sigma}_2]^{-1}(\mu_1 - \mu_2) + \frac{1}{2}\ln\frac{|s\boldsymbol{\Sigma}_1 + (1-s)\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|^s |\boldsymbol{\Sigma}_2|^{1-s}}.$$

2.5 Testy wstępne w analizie dyskryminacyjnej

Niech danych będzie k ($k \geqslant 2$) populacji $\pi_1, \ldots, \pi_k$. W zagadnieniach testów wstępnych populacje te często zwane są "standardami". Ponadto, niech dana będzie jeszcze jedna populacja (π_0), zwana "kandydatem". Test wstępny ma dać odpowiedź na pytanie: czy populacja π_0 jest nową populacją, czy jest to jedna ze znanych "standardowych" populacji. Zatem problem polega na weryfikacji hipotezy zerowej

$$\mathrm{H}_0 :\ \forall_{i \in \{1,\ldots,k\}} :\ \pi_0 \neq \pi_i,$$

przeciwko hipotezie alternatywnej

$$\mathrm{H}_1 :\ \exists_{i \in \{1,\ldots,k\}} :\ \pi_0 = \pi_i.$$

Zagadnienie testów wstępnych po raz pierwszy rozważane było przez Rao (1965). W swojej pracy przyjmował on założenie normalności rozkładów populacji z jednakowymi macierzami kowariancji oraz pełną wiedzę o standardach. W roku 1976 McDonald i inni rozwijali pomysły Rao zakładając, że

informacja o standardach (jedynie dwóch) pochodzi z próby (nie zmieniając założenia o normalności rozkładów populacji z jednakowymi macierzami kowariancji).
Zastosowanie odległości probabilistycznych do rozwiązania problemu testów wstępnych pojawiło się w pracy Bar-Hena (1996), a następnie zostało rozwinięte w pracach Kali i Krzyśki (2003). Zagadnieniu temu poświęcona jest praca [H6].
Przyjmując założenie, że gęstości rozkładów populacji f_i zależą od nieznanego parametru θ_i $(i = 0, 1, \ldots, k)$ mamy $\rho_i = \rho(f_0, f_i) \equiv \rho(\theta_0, \theta_i)$, $i = 1, 2, \ldots, k$. Rozważany układ hipotez można zapisać w postaci:

$$\mathrm{H}_0 : \ \forall_{i\in\{1,\ldots,k\}} : \ \rho_i \geqslant \rho_0,$$

$$\mathrm{H}_1 : \ \exists_{i\in\{1,\ldots,k\}} : \ \rho_i < \rho_0,$$

gdzie ρ_0 $(\rho_0 > 0)$ jest ustalone.
Bar-Hen (1996) pokazał, że przy założeniu warunków regularności wektor losowy $\hat{\boldsymbol{\rho}} = (\hat{\rho}_1, \ldots, \hat{\rho}_k)'$, gdzie $\hat{\rho}_i = \rho(\hat{\theta}_0, \hat{\theta}_i)$ oraz $\hat{\theta}_0$, $\hat{\theta}_i$ są ENW parametrów θ_0, θ_i, ma asymptotycznie k-wymiarowy rozkład normalny z wartością oczekiwaną $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k)'$ i macierzą kowariancji $\boldsymbol{\Sigma} = (\sigma_{ij})$ postaci

$$\sigma_{ij} = \left(\tfrac{\partial\rho_i}{\partial\theta_0}\right)' I^{-1}(\theta_0) \left(\tfrac{\partial\rho_j}{\partial\theta_0}\right), \quad i \neq j,$$

$$\sigma_{ii} = \left(\tfrac{\partial\rho_i}{\partial\theta_0}\right)' I^{-1}(\theta_0) \left(\tfrac{\partial\rho_i}{\partial\theta_0}\right) + \left(\tfrac{\partial\rho_i}{\partial\theta_i}\right)' I^{-1}(\theta_i) \left(\tfrac{\partial\rho_i}{\partial\theta_i}\right),$$

gdzie $I(\theta_i)$ jest macierzą informacji Fishera.
Przyjmując za statystykę testową $\min_{1\leqslant i\leqslant k} \hat{\rho}_i$ wartość krytyczną c dobieramy tak, aby

$$F(c) = \alpha + (1 - \alpha)F(0),$$

gdzie F jest dystrybuantą rozkładu statystyki testowej, a α poziomem istotności testu. Bazując na granicznym rozkładzie wektora losowego $\hat{\boldsymbol{\rho}}$, Kala i Krzyśko (2003) podali postać dystrybuanty F (jako całkę, której krotność zależy od wymiaru przestrzeni parametru).
W pracy [H6] rozważono przypadek populacji o p-wymiarowych rozkładach normalnych z różnymi oraz równymi macierzami kowariancji. W pierwszym przypadku zastosowano odległość Bhattacharyya, natomiast w drugim odległość Mahalanobisa pomiędzy populacjami. Podano jawne postaci, potrzebnych do wyznaczenia wartości krytycznych, pochodnych odległości względem występujących w modelu parametrów. Mówi o tym poniższe twierdzenie.

Twierdzenie 6. *Dla odległości Bhattacharyya mamy:*

$$\frac{\partial \rho_i}{\partial \boldsymbol{\mu}_0} = \frac{1}{4}\boldsymbol{\eta}_i, \quad \frac{\partial \rho_i}{\partial \boldsymbol{\mu}_i} = -\frac{1}{4}\boldsymbol{\eta}_i, \quad i = 1, 2, \ldots, k,$$

$$\frac{\partial \rho_i}{\partial \boldsymbol{\Sigma}_0} = \frac{1}{2}\left(\boldsymbol{V}_i^{-1} - \boldsymbol{\Sigma}_0^{-1} - \frac{1}{4}\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right) - \frac{1}{4}\mathrm{diag}\left(\boldsymbol{V}_i^{-1} - \boldsymbol{\Sigma}_0^{-1} - \frac{1}{4}\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right),$$

$$\frac{\partial \rho_i}{\partial \boldsymbol{\Sigma}_i} = \frac{1}{2}\left(\boldsymbol{V}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} - \frac{1}{4}\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right) - \frac{1}{4}\mathrm{diag}\left(\boldsymbol{V}_i^{-1} - \boldsymbol{\Sigma}_i^{-1} - \frac{1}{4}\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right),$$

gdzie $\boldsymbol{\eta}_i = \boldsymbol{V}_i^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_i)$ *oraz* $\boldsymbol{V}_i = (\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_i)/2$, $i = 1, 2, \ldots, k$.

Ze względu na trudną do praktycznego wykorzystania postać dystrybuanty F rozkładu statystyki testowej, w pracy [H6] wykorzystano do wyznaczania wartości krytycznych metodę symulacyjną bazującą na generowaniu obserwacji z rozkładu granicznego wektora losowego $\hat{\boldsymbol{\rho}}$ i na ich podstawie szacowaniu dystrybuanty F.

2.6 Klasyfikatory jądrowe

W pracach [H2], [H3] i [H4] rozważane były liniowe klasyfikatory maksymalizujące odległości pomiędzy populacjami. W przypadku dwóch populacji π_0, π_1, optymalny liniowy klasyfikator binarny możemy zapisać w postaci

$$d(\boldsymbol{x}) = I(< \boldsymbol{a}, \boldsymbol{x} > \geqslant c),$$

przy czym wektor parametrów $\boldsymbol{a}$ wyznaczamy tak, aby

$$\boldsymbol{a} = \arg\max_{\boldsymbol{w} \in \mathbf{R}^p} J(\boldsymbol{w}),$$

gdzie J jest zadanym kryterium, w naszych rozważaniach jest odległością pomiędzy populacjami.

Niech Φ będzie nieliniowym przekształceniem przestrzeni $\mathbf{R}^p$ w przestrzeń Hilberta z jądrem reprodukującym $\mathcal{H}$. Jak poprzednio w przestrzeni $\mathbf{R}^p$, tak teraz w przestrzeni $\mathcal{H}$ poszukujemy liniowego, binarnego klasyfikatora postaci

$$d(\boldsymbol{x}) = I(< a, \Phi(\boldsymbol{x}) > \geqslant c),$$

gdzie

$$a = \arg\max_{w \in Lin(\mathcal{L}^{\mathcal{H}})} J(w)$$

oraz $\mathcal{L}^{\mathcal{H}}$ jest obrazem przestrzeni próby $\mathcal{L}$ w przekształceniu Φ, a J jest miarą odległości pomiędzy populacjami.

Mamy

$$a = \sum_{j=1}^{n} \alpha_j \Phi(\boldsymbol{x}_j).$$

Zatem

$$< a, \Phi(\boldsymbol{x}) >= \sum_{j=1}^{n} \alpha_j < \Phi(\boldsymbol{x}_j), \Phi(\boldsymbol{x}) > .$$

Zastępując iloczyny skalarne w przestrzeni $\mathcal{H}$ funkcjami jądrowymi otrzymujemy

$$< a, \Phi(\boldsymbol{x}) >= \sum_{j=1}^{n} \alpha_j K(\boldsymbol{x}_j, \boldsymbol{x}).$$

Klasycznymi przykładami funkcji jądrowych (jąder) są jądra wielomianowe $K(\boldsymbol{x}, \boldsymbol{y}) = (1 + \boldsymbol{x}'\boldsymbol{y})^c$ oraz jądra gaussowskie $K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x} - \boldsymbol{y}\|^2/d)$, gdzie c i d są stałymi. Idea zastępowania iloczynów skalarnych przez funkcje jądrowe była wcześniej wielokrotnie wykorzystywana, między innymi przez Vapnika (1995) w metodzie wektorów nośnych, przez Schölkopfa, Smolę i Müllera (1998) w analizie składowych głównych.
Pozostaje jeszcze problem przeniesienia miar odległości pomiędzy populacjami z przestrzeni $\mathbf{R}^p$ do przestrzeni Hilberta $\mathcal{H}$.
Zauważmy, ze dla dowolnego $w \in Lin(\mathcal{L}_n^{\mathcal{H}})$,

$$w = \sum_{j=1}^{n} \omega_j \Phi(\boldsymbol{x}_j).$$

Stąd kryterium optymalności przyjmuje postać

$$\boldsymbol{\alpha} = \arg \max_{\boldsymbol{\omega} \in \mathbf{R}^n} J(\boldsymbol{\omega}).$$

Poniższe twierdzenie podaje postać miary J dla odległości Bhattacharyya.

Twierdzenie 7. *Dla odległości Bhattacharyya mamy:*

$$\begin{aligned} J_B(\boldsymbol{\omega}) = & \frac{1}{4} \frac{\boldsymbol{\omega}'\boldsymbol{M}\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{N}_1\boldsymbol{\omega} + \boldsymbol{\omega}'\boldsymbol{N}_0\boldsymbol{\omega}} + \frac{1}{2} \ln[\frac{1}{2}\boldsymbol{\omega}'\boldsymbol{N}_1\boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{N}_0\boldsymbol{\omega}] \\ & - \frac{1}{4} \ln(\boldsymbol{\omega}'\boldsymbol{N}_1\boldsymbol{\omega}) - \frac{1}{4} \ln(\boldsymbol{\omega}'\boldsymbol{N}_0\boldsymbol{\omega}), \end{aligned}$$

gdzie

$$\boldsymbol{M} = (\boldsymbol{M}_1 - \boldsymbol{M}_0)(\boldsymbol{M}_1 - \boldsymbol{M}_0)', \quad (M_i)_j = \frac{1}{n_i} \sum_{k=1}^{n_i} K(\boldsymbol{x}_j, \boldsymbol{x}_k^i)$$

oraz

$$\boldsymbol{N}_i = \frac{1}{n_i - 1} \boldsymbol{K}_i (\boldsymbol{I} - \frac{1}{n_i} \boldsymbol{1}\boldsymbol{1}') \boldsymbol{K}_i', \quad (K_i)_{kl} = K(\boldsymbol{x}_k, \boldsymbol{x}_l^i), \quad i = 0, 1.$$

W ten sposób postawiony w przestrzeni Hilberta problem został zredukowany do problemu wyznaczenia wektora $\boldsymbol{\alpha}$ w przestrzeni $\mathbf{R}^n$. Niestety występujące w kryterium J_B macierze $\boldsymbol{N}_0$ oraz $\boldsymbol{N}_1$ nie są dodatnio określone. Aby móc wykorzystać algorytm przedstawiony z prac [H2] i [H3] dokonujemy ich regularyzacji, tzn. zastępujemy je przez macierze $\boldsymbol{N}_0^\star = \boldsymbol{N}_0 + \varepsilon_0 \boldsymbol{I}$ i $\boldsymbol{N}_1^\star = \boldsymbol{N}_1 + \varepsilon_1 \boldsymbol{I}$ odpowiednio. Stałe ε_0 i ε_1 dobieramy tak, aby były możliwie małe lecz zapewniały dodatnią określoność macierzy $\boldsymbol{N}_0^\star$ i $\boldsymbol{N}_1^\star$.
Metoda zaproponowana w pracy [H8] pozwala na połączenie prostoty klasyfikatorów liniowych, z nieliniowymi powierzchniami rozdzielającymi populacje. Jak pokazują przykłady zamieszczone w pracy [H8] pozwala to na znaczną redukcję liczby błędnych klasyfikacji.

7. Pozostały dorobek naukowy:

Mój pozostały dorobek naukowy, po uzyskaniu stopnia doktora, składa się z dwóch książek, 24 prac naukowych oraz jednej pracy popularno-naukowej.

A. Książki:

1. M. Krzyśko, W. Wołyński, T. Górecki, M. Skorzybut, Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości. WNT, Warszawa 2008.
2. W. Wołyński, Prawdopodobieństwo i statystyka. Zadania z egzaminów dla Aktuariuszy z rozwiązaniami 2003-2007. Wydawnictwo Naukowe UAM, Poznań 2008.

B. Prace naukowe:

1. M. Krzyśko, W. Wołyński, *Statistical group classification rules for the multivariate Student t-distribution*, Random Operators and Stochastic Equations **1** (1993), 361-367.
2. M. Krzyśko, W. Wołyński, *UMVU estimators in statistical group classification rules*, Random Operators and Stochastic Equations **2** (1994), 141-152.
3. W. Wołyński, *A comparison of an estimative and predictive method of the group statistical discrimination*, Random Operators and Stochastic Equations **3** (1995), 75-86.
4. W. Wołyński, *Unbiased classification using multiple observations*, Discussiones Mathematicae. Algebra and Stochastic Methods, **16** (1996), 135-143.

5. W. Jassem, M. Krzyśko, W. Wołyński, *Normalisation of Polish Vowel Spectra*, The Phonetician **81** (2000), 23-31.

6. M. Krzyśko, W. Wołyński, *Classification into two populations for time dependent observations*, Acta Universitatis Lodziensis. Folia Oeconomica **152** (2000), 7-20.

7. Z. Młynarek, W. Tschuschke, W. Wołyński, *Statistical evaluation of different in situ tests in post-flotation sediments*, Proceedings of the International Conference on In Situ Measurement of Soil Properties and Case Histories, Bali, Indonesia 2001, 679-683.

8. Z. Młynarek, W. Tschuschke, W. Wołyński, *The CTPU classification chart for the post-flotation sediments based on statistical criteria*, Proceedings of the fifteenth International Conference on Soil Mechanics and Geotechnical Engineering, Istambul 2001, A.A.Balkema Publishers, 459-462.

9. W. Jassem, M. Krzyśko, W. Wołyński, *Reduction of speaker-related variability in Polish vowel spectra*, Journal of the International Phonetic Association **31** (2001), 187-202.

10. J. Wierzbicki, W. Wołyński, *Statistical model determining coefficient of earth pressure at rest by means of static penetration, case of noncohesive soil*, Studia Geotechnica et Mechanica **24** (2002), 27-38.

11. A. Niedzielski, J. Wierzbicki, M. Waliński, W. Wołyński, *The quality control of post flotation reservoir dam by determination of relative compaction index in various methods*, Proceedings ISC-2 on Geotechnical and Geophysical Site Characterization, Viana da Fonseca & Mayne (eds.), 2004 Millpress, Rotterdam, 1371-1375.

12. Z. Młynarek, J. Wierzbicki, W. Wołyński, *Use of cluster method for in situ tests*, Studia Geotechnica et Mechanica, **27** (2005), 15-27.

13. Z. Młynarek, J. Wierzbicki, W. Wołyński, *Use of interpolation methods for geotechnical profiling*, Studia Geotechnica et Mechanica, **27** (2005), 5-13.

14. J. Wierzbicki, W. Wołyński, *Ocena zagęszczenia odpadów poflotacyjnych z uwzględnieniem założonego prawdopodobieństwa awarii*, Roczniki Akademii Rolniczej w Poznaniu, Melior. Inż. Środ. **26** (2005), 503-509.

15. Z. Młynarek, W. Tschuschke, J. Wierzbicki, W. Wołyński, *Statistical criteria of determination of homogenous geotechnical layers*,

Proceedings of the 16th International Conference on Soil Mechanics and Geotechnical Engineering, Osaka 2005, Millpress, Rotterdam, 725-728.

16. T. Caliński, M. Krzyśko, W. Wołyński, *A comparison of some tests for determining the number of nonzero canonical correlations*, Communications in Statistics - Simulation and Computation, **35** (2006), 727-749. IF: 0.482.

17. Z. Młynarek, J. Wierzbicki, W. Wołyński, *An approach to 3D subsoil model based on CPTU results*, Proceedings of the 14th European Conference on Soil Mechanics and Geotechnical Engineering, Madrid 2007, Millpress, Rotterdam, 1721-1726.

18. Z. Młynarek, J. Wierzbicki, W. Wołyński, *Efficiency of selected statistical criteria in determination of geotechnical parameters from CPTU*, Studia Geotechnica et Mechanica, 29 (2007), 137-149.

19. A. Nosalewicz, D. Szynal, W. Wołyński, *Empirical study of tests for exponentiality*, Transactions of the XXVII International Seminar on Stability Problems for Stochastic Models, Prof. Zeev Volkovich (ed.), Karmiel (Israel) 2007, 157-164.

20. Z. Młynarek, J. Wierzbicki, W. Wołyński, W. Tschuschke, *Assessment of efficiency of different cluster analysis methods for evaluation of a stratigraphy of strongly laminated subsoil*, Proc. of the 12th Int. Conference of International Association for Computer Methods and Advances in Geomechanics. Goa, India 2008, 1291-1299.

21. M. Krzyśko, S. Pluta, M. Skorzybut, W. Wołyński, *Analysis of multivariate repeated measures data*, Colloquium Biometricum **40** (2010), 117-133.

22. W. Wołyński, *Effectiveness of decomposition algorithms for multiclass classification problems*, Acta Universitatis Lodziensis. Folia Oeconomica **235** (2010), 205-213.

23. M. Krzyśko, M. Skorzybut, W. Wołyński, *Classifiers for doubly multivariate data*, Discussiones Mathematicae. Probability and Statistics **31** (2011), 5-27.

24. D. Szynal, W. Wołyński, *Goodness-of-fit tests for exponentiality and Rayleigh distribution*, IJPAM **78** (2012), 751-772.

C. Praca popularno-naukowa:

1. Z. Młynarek, W. Tschuschke, J. Wierzbicki, W. Wołyński, *Wykorzystanie statystycznej analizy danych do wydzielania geotechnicznych warstw podłoża budowlanego*, Geoinżynieria i tunelowanie **2** (2005), 14-17.

Książki

Książka "Systemy uczące się. Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości" wydana w 2008 roku w Wydawnictwach Naukowo Technicznych zawiera omówienie wielu technik klasyfikacyjnych wykorzystujących metody statystyczne. Składa się z dwóch części. W pierwszej przedstawiono systemy uczenia się pod nadzorem. Omówiono w niej różnorodne metody klasyfikacji: probabilistyczne, regresyjne, metodę wektorów nośnych, metodę najbliższego sąsiada, drzewa regresyjne i sieci neuronowe. W części tej omówione zostały również zagadnienia estymacji błędów, techniki dekompozycji i łączenia klasyfikatorów oraz metody ich wzmacniania.
W części drugiej przedstawiono systemy uczenia się bez nadzoru takie, jak analiza składowych głównych, analiza skupień, skalowanie wielowymiarowe oraz analizę korespondencji.

Na prośbę wielu studentów Wydziału Matematyki i Informatyki UAM oraz Uniwersytetu Ekonomicznego w Poznaniu przygotowujących się do egzaminu dla aktuariuszy, przygotowałem i opublikowałem w Wydawnictwie Naukowym UAM, pełne rozwiązania 12 zestawów egzaminacyjnych z zakresu rachunku prawdopodobieństwa i statystyki. Wiem, że moje książka pomogła w przygotowaniach do egzaminu. Same zaś zadania (z reguły ciekawe i niebanalne) można również wykorzystać podczas prowadzenia zajęć z rachunku prawdopodobieństwa i statystyki dla studentów zarówno matematyki, jak i wielu kierunków ekonomicznych.

Prace naukowe

Prace [1], [2] i [3] opublikowane w latach 1993-1995 w czasopiśmie "Random Operators and Stochastic Equations" dotyczą zagadnień klasyfikacji grup obiektów. Tematyką tą interesowałem się od początku swojej działalności naukowej, była ona również tematem mojej pracy doktorskiej.
Głównym wynikiem pracy [1] było podanie postaci ENMW funkcji klasyfikującej w przestrzeni statystyk dostatecznych przy założeniu, że obserwacje pochodzą z populacji o p-wymiarowym rozkładzie t-Studenta. Wynik ten zawiera poniższe twierdzenie:

Twierdzenie 8. *Niech q_i będzie prawdopodobieństwem a priori przynależności grupy $\boldsymbol{x}_0$ do populacji π_i, $N_0 > p$, $N_i > N_0 + p$, $i = 1, 2, \ldots, k$. Ponadto niech*

$$\bar{x}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}, \quad A_i = \sum_{j=1}^{N_i}(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)', \quad i = 0, 1, \ldots, k.$$

Wtedy ENMW funkcji klasyfikującej $u_i(\bar{x}_0, A_0)$ ma postać:

$$\begin{aligned}\hat{u}_i(\bar{x}_0, A_0) &= q_i c_i \mid A_0 \mid^{\frac{N_0-p-2}{2}} \mid A_i \mid^{-\frac{N_i-p-2}{2}} \\ &\times \left[\psi\Big(A_i - A_0 - \frac{N_0 N_i}{N_i - N_0}(\bar{x}_0 - \bar{x}_i)(\bar{x}_0 - \bar{x}_i)'\Big)\right]^{\frac{N_i-N_0-p-2}{2}},\end{aligned}$$

gdzie

$$c_i = \Big[\frac{N_i N_0}{\pi(N_i - N_0)}\Big]^{\frac{p}{2}} \frac{\Gamma_p\Big(\frac{N_i-1}{2}\Big)}{\Gamma_p\Big(\frac{N_0-1}{2}\Big)\Gamma_p\Big(\frac{N_i-N_0-1}{2}\Big)}, \quad i = 1, 2, \ldots, k,$$

oraz

$$\psi(A) = \begin{cases} \mid A \mid & gdy \;\; A > 0, \\ 0 & poza \; tym. \end{cases}$$

W pracy [2] rozważano to samo zagadnienie co w pracy [1] tyle tylko, że w p-wymiarowym modelu normalnym. Poza analogicznym do twierdzenia 8 wynikiem podającym ENMW funkcji klasyfikującej w przestrzeni statystyk dostatecznych, udowodniono również twierdzenie mówiące o tym, że estymator ten jest zgodny. Ponadto, w pracy tej, podano postać średniokwadratowego błędu całkowego (MISE) tego estymatora. Wynik ten podaje twierdzenie 9.

Twierdzenie 9. *Dla ENMW funkcji klasyfikującej u_i w przestrzeni statystyk dostatecznych mamy*

$$MISE(\hat{u}_i(\bar{x}_0, A_0)) = q_i c_p \mid \boldsymbol{\Sigma} \mid^{-\frac{p+2}{2}} \Big(2^{\frac{p(2N_0-p-2)}{2}}\gamma_p - 1\Big),$$

gdzie

$$c_p = \Big(\frac{N_0}{\pi}\Big)^{\frac{p}{2}} 2^{-pN_0} \frac{\Gamma_p\Big(\frac{2N_0-p-3}{2}\Big)}{\Gamma_p^2\Big(\frac{N_0-1}{2}\Big)}$$

oraz

$$\gamma_p = \Big(\frac{N_i}{N_i - N_0}\Big)^{\frac{p}{2}} \frac{\Gamma_p\Big(\frac{N_i-1}{2}\Big)\Gamma_p\Big(\frac{N_i-p-3}{2}\Big)\Gamma_p\Big(\frac{2N_i-2N_0-p-3}{2}\Big)}{\Gamma_p^2\Big(\frac{N_i-N_0-1}{2}\Big)\Gamma_p\Big(\frac{2N_i-2p-5}{2}\Big)}, \quad i = 1, \ldots, k.$$

Jako wniosek z tego twierdzenia otrzymujemy, że

$$MISE(\hat{u}_i(\bar{x}_0, A_0)) = q_i c_p \mid \mathbf{\Sigma} \mid^{-\frac{p+2}{2}} \left\{ \frac{pN_0}{2N_i} + \frac{p[2N_0(p+1)+p+2]}{8N_i} + O(N_i^{-2}) \right\}.$$

Ostatnią z cyklu trzech prac poświęconych estymacji funkcji klasyfikujących w przestrzeni statystyk dostatecznych jest praca [3]. Zastosowano w niej podejście bayesowskie do rozwiązania problemu optymalnej estymacji. Zakładając p-wymiarowy model normalny oraz przyjmując rozkład a priori parametrów $\boldsymbol{\mu}_i$ i $\mathbf{\Sigma}_i$ postaci

$$g(\boldsymbol{\mu}_i, \mathbf{\Sigma}_i^{-1}) \propto |\mathbf{\Sigma}_i|^{(p+1)/2)}, \; i = 1, \ldots k,$$

udowodniono następujące twierdzenie:

Twierdzenie 10. *Niech $N_i > p$. Wtedy estymator bayesowski funkcji klasyfikującej $u_i(\bar{x}_0, A_0)$ ma postać:*

$$\begin{aligned} \hat{u}_i(\bar{x}_0, A_0) = {} & q_i c_i \mid A_0 \mid^{\frac{N_0-p-2}{2}} \mid A_i \mid^{\frac{N_i-1}{2}} \\ & \times \mid A_0 + A_i + \frac{N_0 N_i}{N_i + N_0} (\bar{x}_0 - \bar{x}_i)(\bar{x}_0 - \bar{x}_i)' \mid^{-\frac{N_0+N_i-1}{2}}, \end{aligned}$$

gdzie

$$c_i = \left[\frac{N_0 N_i}{\pi(N_i + N_0)} \right]^{\frac{p}{2}} \frac{\Gamma_p\left(\frac{N_0+N_i-1}{2}\right)}{\Gamma_p\left(\frac{N_0-1}{2}\right)\Gamma_p\left(\frac{N_i-1}{2}\right)}, \quad i = 1, 2, \ldots, k.$$

Dodatkowo, w pracy [3] podano wyniki symulacyjnego porównania jakości estymatorów funkcji klasyfikującej. Do porównań użyto estymatory częstościowe, nieobciążone o minimalnej wariancji oraz bayesowskie.

W 1995 roku uczestniczyłem (jako stypendysta) w odbywających się w Milton Keynes (Anglia) European Courses in Advanced Statistics, których tematyka dotyczyła między innymi analizy danych pochodzących z doświadczeń z powtarzanymi pomiarami. Plonem tego wyjazdu była praca [4] dotycząca estymacji funkcji dyskryminacyjnych w modelu z powtarzanymi pomiarami. Oznaczając przez X_{ijk}, k-ty $(p \times 1)$ wektor obserwacji j-tego obiektu w populacji i-tej $(i = 1, 2;\; j = 1, 2, \ldots, n_i;\; k = 1, 2, \ldots, n)$ rozważany był następujący model

$$X_{ijk} = \mu_i + I_{ij} + E_{ijk}$$

gdzie

(i) μ_i jest stałe,

(ii) $I_{ij} \sim iid\ N_p(0, \Omega_i)$,

(iii) $E_{ijk} \sim iid\ N_p(0, \Sigma_i)$,

(iv) I_{ij} i E_{ijk} są niezależne.

Ponadto, niech $\boldsymbol{X}_0 = (X_{01}, X_{02}, \ldots, X_{0n})$, $n > 1$, będzie nową (klasyfikowaną) obserwacją, przy czym kolumny odpowiadają wielokrotnym pomiarom. Model powyższy, w zastosowaniu do problemów klasyfikacyjnych, rozważany był w pracach Choi (1972), Gupty (1986) oraz Gupty i Logana (1990, 1993). Głównymi wynikami pracy są twierdzenia podające postaci estymatorów nieobciążonych o minimalnej wariancji funkcji gęstości wektora $\boldsymbol{X}_0$ oraz funkcji dyskryminacyjnej. Ponadto dokonano symulacyjnego porównania estymatora nieobciążonego o minimalnej wariancji z estymatorem częstościowym oraz podanym przez Guptę i Logana (1993) estymatorem bayesowskim.

Prace [5] i [9] są efektem współpracy z prof. drem hab. Wiktorem Jassemem z Pracowni Fonetyki Akustycznej Instytutu Podstawowych Problemów Techniki PAN. Obydwie prace dotyczą zagadnienia automatycznego rozpoznawania mowy, w szczególności w pracy [5] rozpoznawaniu podlega mówiąca osoba natomiast w pracy [9] rozpoznawana jest wypowiadana polska samogłoska. Do identyfikacji wykorzystano znormalizowane formanty samogłoskowe. W pracy [5] jako metodę klasyfikacyjną przyjęto klasyczną metodę liniową Fishera (LDA) skupiając się głównie na zbadaniu wpływu normalizacji na jakość rozpoznawania. W pracy [9] zastosowano do klasyfikacji, opracowaną przez Krzyśkę (1999) metodę liniową bazującą na maksymalizacji pola powierzchni pod krzywą ROC.

W pracy [6] zaproponowano użycie krzywej ROC do badania jakości liniowych procedur klasyfikacyjnych ilustrując tę technikę przykładem klasyfikacji procesów autoregresyjnych rzędu drugiego.

Jak już wspomniałem w rozdziale 2.4. w 1999 roku nawiązałem współpracę z pracownikami Katedry Geotechniki Uniwersytetu Przyrodniczego w Poznaniu. Plonem tej współpracy jest 12 prac ([7], [8], [10]-[15], [17], [18], [20] oraz praca popularno-naukowa [1]) opublikowanych począwszy od 2001 roku. Uzyskane wyniki były wielokrotnie prezentowane na prestiżowych międzynarodowych konferencjach z zakresu geotechniki oraz mechaniki gruntów odbywających się w Turcji, Indonezji, Portugalii, Japonii, Hiszpanii oraz Indii. Część z nich została również opublikowana w polskim czasopiśmie "Studia Geotechnica et Mechanica". W pracach tych zastosowano metody statystyki

matematycznej (głównie wielowymiarowej) do analizy wyników badań gruntów metodą sondowania statycznego (CPT - Cone Penetration Test) polegającą na wciskaniu w podłoże stożkowo zakończonego elementu pomiarowego. Szczegółowy opis tej metody zawiera praca Lunne, Młynarka i Tschuschke (1995) oraz monografia Lunne, Powella i Robertsona (1997). Analiza tak uzyskanych danych obejmowała wiele różnych technik statystycznych, takich jak: analiza dyskryminacyjna, zastosowana do klasyfikacji gruntów poflotacyjnych (praca [8]); regresja prosta i wielokrotna, zastosowana w analizie zależności pomiędzy wielkościami uzyskanymi z sondowania statycznego, a innymi zmiennymi (np. niedrenowaną wytrzymałością na ścinanie, współczynnikiem parcia spoczynkowego) wykorzystywanymi do analizy gruntów (prace [7] i [10]); analiza wariancji, wykorzystana do porównania metod pomiarowych (praca [11]); analiza skupień, wykorzystana do wyznaczania profili podłoża (prace [12], [15], [18], [20] oraz praca popularno-naukowa [1]); kriging oraz analiza korelacji kanonicznych, zastosowana do tworzenia map geotechnicznych gruntów (prace [13] i [17]); teoria niezawodności, wykorzystana do szacowania ryzyka awarii budowli ziemnych (praca [14]).

Jestem także współautorem raportu przygotowanego przez Katedrę Geotechniki Uniwersytetu Przyrodniczego w Poznaniu oraz Norweski Instytut Geotechniki w sprawie oceny jakości pomiarów wykonywanych sondami statycznymi czołowych w świecie producentów tych urządzeń.

Cały czas poprawiamy, uzupełniamy oraz tworzymy nowe procedury. Impulsem do dalszej pracy jest stały rozwój urządzeń i technik pomiarowych wykorzystywanych przy badaniu gruntów.

Praca [16] dotyczy zagadnienia wymiarowości w analizie kanonicznej. Impulsem do jej powstania było opracowanie przez Calińskiego i Krzyśkę (2005) tzw. zamkniętej procedury testowej. Zawiera ona wyniki badań porównawczych siedmiu procedur testowych oraz kryterium informacyjnego Akaike.

Prace [19] i [24] są plonem mojej współpracy z profesorem Dominikiem Szynalem z Instytutu Matematyki Uniwersytetu Marii Curie-Skłodowskiej w Lublinie. Omawiane są w nich zagadnienia testowania wykładniczości rozkładu populacji (również zgodności rozkładu populacji z rozkładem Rayleigh'a). Analizowano w nich procedury testowe oparte na statystykach rekordowych. Za pomocą badań symulacyjnych wyznaczone zostały tzw. omnibus tests. Porównano również moce zaproponowanych nowych testów z mocami 8 innych znanych procedur służących do testowania wykładniczości. Szczególną uwagę zwrócono na przypadek małych prób (n=10, n=20) dla których trudno uzyskać testy o zadawalających mocach.
W mojej opinii szczególnie wartościowym wynikiem badań są zaproponowane, proste i jednocześnie relatywnie mocne procedury testowania wykładni-

czości rozkładu populacji. W zaproponowanych procedurach statystyki testowe mają następujące postaci:

$$T_1(\boldsymbol{x}) = \frac{16n}{16-5\sqrt{\pi}}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\bar{x}}\right)^{1/2} - \frac{\sqrt{\pi}}{2}\right]^2$$

oraz

$$T_2(\boldsymbol{x}) = \frac{5n(n-1)(n-2)}{4n^2-3n-7}\left[\frac{9}{n(n-1)(n-2)}\sum_{i=1}^{n-2}(n-i-1)(n-i)\frac{x_{i:n}}{\bar{x}} - 1\right]^2.$$

Statystyki te mają granicznie (przy prawdziwości hipotezy zerowej) rozkład chi-kwadrat z jednym stopniem swobody. W naszych pracach, ze względu na małe liczebności prób, wykorzystywaliśmy empiryczne wartości krytyczne, co znacząco poprawiało moce tych testów.
Dobre wyniki uzyskiwane w przypadku rozkładów wykładniczego oraz Rayleigh'a, skłoniły nas do kontynuowania badań nad procedurami testowymi opartymi na statystykach rekordowych w zastosowaniu do innych rozkładów. W szczególności planujemy zastosować tego typu procedury do testowania normalności rozkładu.

Prace [21] i [23] dotyczą zagadnień analizy danych podwójnie wielowymiarowych. Dane tego typu otrzymujemy w wyniku T krotnego powtarzania pomiarów p cech danej jednostki doświadczalnej. Niech x_{it} oznacza obserwację i-tej cechy w czasie t ($i = 1,\ldots,p$, $t = 1,\ldots,T$.). Ponadto, niech $\boldsymbol{X} = \mathrm{vec}(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_p)$, gdzie $\boldsymbol{X}_i = (x_{i1},\ldots,x_{iT})$. Dalej zakładamy, że wektor $\boldsymbol{X}$ ma pT wymiarowy rozkład normalny z wektorem wartości oczekiwanych $\boldsymbol{\mu}$ oraz macierzą kowariancji $\boldsymbol{\Omega}$ postaci $\boldsymbol{\Omega} = \boldsymbol{V}\otimes\boldsymbol{\Sigma}$. Aby zastosować klasyczny klasyfikator bayesowski do tego typu danych potrzebne są estymatory parametrów $\boldsymbol{\mu}$ oraz $\boldsymbol{\Omega} = \boldsymbol{V}\otimes\boldsymbol{\Sigma}$ w każdej z klas. W pracy [23] zaproponowano następujący algorytm pozwalający na wyznaczenie estymatorów największej wiarogodności macierzy $\boldsymbol{V}$ oraz $\boldsymbol{\Sigma}$.

Krok 1. Za początkowe oszacowanie macierzy $\boldsymbol{\Sigma}$ przyjmujemy macierz:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{nT}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

Krok 2. Obliczamy oszacowanie macierzy $\boldsymbol{V}$ postaci:

$$\hat{\boldsymbol{V}} = \frac{1}{np}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

Krok 3. Obliczamy oszacowanie macierzy $\mathbf{\Sigma}$ postaci:

$$\hat{\mathbf{\Sigma}} = \frac{1}{nT}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})\hat{\boldsymbol{V}}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})'.$$

Krok 4. Powtarzamy kroki 2 i 3 aż do uzyskania zbieżności.

Algorytm ten może być wykorzystywany w przypadku bardzo małych prób. Musi być jedynie spełniony warunek $n > \max(p, T)$. W celu zmniejszenia liczby parametrów modelu, a tym samym zmniejszenia liczby potrzebnych obserwacji przyjmowano dotychczas dodatkowe założenia dotyczące macierzy $\boldsymbol{V}$ (Roy, Khattree (2005, 2008)), np. zakładano model autoregresji rzędu pierwszego. W pracy [23] podane są również zmodyfikowane wersje powyższego algorytmu zakładające dodatkowe założenia o macierzy $\boldsymbol{V}$. Z zamieszczonych w niej przykładów wynika, że używanie bardziej ogólnego modelu prowadzi do redukcji błędów w bayesowskich procedurach klasyfikacyjnych.

Praca [21] pokazuje możliwości zastosowania oraz jakość procedur dla danych podwójnie wielowymiarowych na konkretnym przykładzie. Wykorzystano w tym celu dane pochodzące z wieloletniego doświadczenia z czarną porzeczką przeprowadzonego w Instytucie Ogrodnictwa w Skierniewicach.

Zagadnienie łączenia informacji pochodzących z klasyfikatorów binarnych jest szczególnie ważne wtedy, gdy procedura klasyfikacyjna nie daje się bezpośrednio rozszerzyć na przypadek wieloklasowy. Jak już wspomniałem w rozdziale 2.3 nawet wtedy, gdy istnieje rozszerzenie wieloklasowe, dekompozycja zagadnienia wieloklasowego na szereg zagadnień binarnych, a następnie połączenie informacji z takich binarnych klasyfikatorów daje często lepsze wyniki niż bezpośrednie użycie procedury wieloklasowej. Praca [22] poświęcona jest badaniu efektywności takiego podejścia. Wyniki badań zamieszczonych w tej pracy wskazują, że zwłaszcza w przypadku "niestabilnych" procedur klasyfikacyjnych takich jak drzewa klasyfikacyjne czy sieci neuronowe dekompozycja zapewnia znaczną poprawę jakości klasyfikacji. Zatem procedury dekompozycyjne można zaliczyć do technik wzmacniania klasyfikatorów.
Zagadnienia łączenia informacji, mierzenia zgodności informacji pochodzących z różnych źródeł są nadal mocno badane. Cały czas prowadzę badania związane z zagadnieniami wzmacniania klasyfikatorów.

8. Projekty badawcze:

1. Wykonawca w projekcie badawczym, numer 210809101, realizacja 1992-1994. Tytuł: Wielowymiarowa analiza statystyczna: metody i zastosowania. Kierownik: prof. dr hab. Mirosław Krzyśko (UAM, Poznań).

2. Wykonawca w projekcie finansowanym przez Ministerstwo Nauki i Informatyzacji, numer 2P04E01929, realizacja 2005-2007. Tytuł: Analiza stopnia przekonsolidowania wybranych osadów kenozoicznych z wykorzystaniem metod in situ. Kierownik: dr hab. Jędrzej Wierzbicki (Uniwersytet Przyrodniczy w Poznaniu).

9. Nagrody naukowe:

Nagroda Dziekana Wydziału Matematyki i Informatyki UAM za działalność naukową w roku 2005.

10. Staże naukowe:

Uniwersytet w Permie, Rosja; staż naukowy; 1991.
The Open University, Anglia; stypendium - European Courses in Advanced Statistics; 1995.

11. Referaty wygłoszone na konferencjach naukowych:

1. XX Konferencja "Statystyka Matematyczna", Wisła, 5-9 grudnia 1994. Referat: *Klasyfikacja grup obiektów metodami odległościowymi w przestrzeni statystyk dostatecznych.*

2. XXI Konferencja "Statystyka Matematyczna", Wisła, 4-8 grudnia 1995. Referat: *Estymacja funkcji dyskryminacyjnych w modelu z powtarzanymi pomiarami.*

3. 6th International Conference on Mathematical Statistics, Jachranka, 17-21 czerwca 1996. Referat: *Linear discriminant functions for stationary time series* (współautor: M. Krzyśko).

4. XXIII Konferencja "Statystyka Matematyczna", Wisła, 8-12 grudnia 1997. Referat: *Optymalna liczebność próby w klasyfikacji grupowej.*

5. International Biometrical Colloquium in Honour of Tadeusz Caliński, Inowrocław, 14-17 września 1998. Referat: *Optimal sample size in the group classification problem.*

6. XXX Międzynarodowe Colloquium Biometryczne, Wigry, 17-21 września 2000. Referat: *Linear discrimination rules for the multivariate Student t-distribution.*

7. The Ninth International Workshop in Mathematics, Gronów, 24-28 września 2001. Referat: *Combining multiple classification rules* (współautor: M. Krzyśko).

8. International Seminar - CPT, CPTU methods for identification of geotechnical parameters of soil and mine tailings, Gronów, 8-9 października 2001. Zaproszony referat: *Statistical classification systems for evaluation of the data.*

9. XXVII Konferencja "Statystyka Matematyczna", Wisła, 3-7 grudnia 2001. Referat: *Łączenie reguł klasyfikacyjnych* (współautor: M. Krzyśko).

10. XXI Konferncja "Wielowymiarowa analiza statystyczna", Łódź, 4-6 listopada 2002. Referat zaproszony: *Łączone reguły klasyfikacyjne* (współautor: M. Krzyśko).

11. XXVIII Konferencja "Statystyka Matematyczna", Wisła, 2-6 grudnia 2002. Referaty: *Porównanie testów wymiarowości w analizie kanonicznej* (współautorzy: T. Caliński, M. Krzyśko) oraz *O pewnym teście poprzedzającym analizę dyskryminacyjną* (współautorzy: R. Kala, M. Krzyśko).

12. Statistical Inference in Linear Models, Będlewo, 21-27 sierpnia 2003. Referat zaproszony: *A comparison of some tests for dimensionality in the canonical correlation analysis* (współautorzy: T. Caliński, M. Krzyśko).

13. 2nd International Seminar - Interpretation of in situ tests and sample disturbance of clays, Baranowo, 23-25 maja 2004. Referat zaproszony: *Use of cluster method for in situ tests* (współautorzy: Z. Młynarek, J. Wierzbicki).

14. 13th International Workshop on Matrices and Statistics, Będlewo, 18-21 sierpnia 2004. Referat: *How to avoid overinterpretation of the results of statistical analyses in medical research* (współautor: J. Hauke).

15. XXX Konferencja "Statystyka Matematyczna", Wisła, 6-11 grudnia 2004. Referat: *Metody korekcji błędów w procedurach klasyfikacyjnych opartych na metodach łączenia klas w pary* (współautor: M. Krzyśko).

16. XXIV Konferencja "Wielowymiarowa Analiza Statystyczna", Łódź, 7-9 listopada 2005. Referat: *Klasyfikacja poprzez łączenie grup w pary* (współautor: M. Krzyśko).

17. XXXI Konferencja "Statystyka Matematyczna", Wisła, 5-9 grudnia 2005. Referat: *Metody szacowania wymiarowości w analizie kanonicznej.*

18. Ogólnopolska Konferencja Naukowa: Statystyka w praktyce społeczno-gospodarczej, Wrocław, 19-21 czerwca 2006. Referat: *Podział populacji na ich naturalne skupienia* (współautor: M. Krzyśko).

19. 3rd International Seminar - Soil design parameters from in situ and laboratory tests, Baranowo, 25-27 września 2006. Referat: *Efficiency of selected statistical criteria in determination of geotechnical parameters from CPTU* (współautorzy: Z. Młynarek, J. Wierzbicki).

20. XXVII Konferencja "Wielowymiarowa Analiza Statystyczna", Łódz, 3-5 listopada 2008. Referat: *Efektywność algorytmów dekompozycyjnych wieloklasowych zagadnień klasyfikacyjnych.*

21. XVIII Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS, Klasyfikacja i analiza danych - teoria i zastosowania, Międzyzdroje, 15-18 września 2009. Referat: *Jądrowe klasyfikatory liniowe.*

22. XXXV Konferencja "Statystyka Matematyczna", Wisła, 7-11 grudnia 2009. Referat: *Jądrowe klasyfikatory liniowe.*

23. Second bilateral German-Polish symposium on data analysis and its applications, Kraków, 14-16 kwietnia, 2011. Referat: *Kernel linear discriminant functions.*

12. Cytowane prace:

R.A. Abusiev, Ya.P. Lumelsky (1980), Unbiased estimators and classification problems for multivariate normal populations, *Teoriya Verojatnostej i Ee Primeneniya*, **25**, 381–389.

T.W. Anderson, R.R. Bahadur (1962), Classification into two multivariate normal distributions with different covariance matrices, *Ann. Math. Statist.* **33**, 420–431.

A. Bar-Hen (1996), A preliminary test in discriminant analysis, *J. Multivar. Anal.* **57** 266—276.

A. Bhattacharyya (1943), On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.* **35**, 99–109.

T. Caliński, M. Krzyśko (2005), A closed testing procedure for canonical correlations, *Commun. Statist. – Theor. Meth.* **34** 1105–1116.

G. Chaudhuri, J.D. Borwankar, P.R.K. Rao (1991a), Bhattacharyya distance - based linear discrimination, *J. Indian Statist. Assoc.* **29**, 47–56.

G. Chaudhuri, J.D. Borwankar, P.R.K. Rao (1991b), Bhattacharyya distance based discriminant function for stationary time series, *Comm. Statist. Theory Methods* **20**, 2195–2205.

H. Chernoff (1952), A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statist.* **23**, 493–507.

S.C. Choi (1972), Classification of multiple observed date, *Biometrische Zeitschrift* **14**, 8–11.

R.A. Fisher (1936), The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7**, 179–188.

A.K. Gupta (1986), On a classification rule for multiple measurements, *Comp. & Maths. with Appls.* **12(A)**, 301–308.

A.K. Gupta, T.P. Logan (1990), On a multiple observations model in discriminant analysis, *J. Statist. Comput. Simul.* **34**, 119–132.

A.K. Gupta, T.P. Logan (1993), Bayesian discrimination using multiple observations, *Commun. Statist. – Theory Meth.* **22(6)**, 1735–1754.

T. Hastie, R. Tibshirani (1996), Classification by pairwise coupling, Technical Report, Stanford University and University of Toronto.

T. Hastie, R. Tibshirani (1998), Classification by pairwise coupling, *Ann. Statist.* **26**, 451–471.

R.L. Iman, J.M. Davenport (1980), Approximations of the critical region of the Friedman statistic. *Communications in Statistics. Theory and Methods*, **A9**, 571–595.

J. Jelonek, J. Stefanowski (1998), Experiments on solving multiclass learning problems by n^2–classifier. In *Proceedings of the 10th European Conference on Machine Learning (ECML-98)*, 172–177, Chemnitz, Germany, 1998.

R. Kala, M. Krzyśko (2003), An extension of Bar-Hen's preliminary test procedure, *J. Multivar. Anal.* **84**, 410-–412.

M. Krzyśko (1999), Linear discriminant functions which maximize the area under the ROC curve. *Discussiones Mathematicae: Algebra and Stochastic Methods* **19**, 335–344.

S. Kullback, A. Leibler (1951), On information and sufficiency, *Ann. Math. Statist.* **22**, 79–86.

T. Lunne, Z. Młynarek, W. Tschuschke (1995), Application of cone penetration test for evaluation of geotechnical parameters of post flotation sediments, Proc. of the Int. Symp. on Cone Penetration Testing CPT'95, Linkoping, Sweden, 329–336.

T. Lunne, J.J.M. Powell, P.K. Robertson (1997), *Cone Penetration Testing in Geotechnical Practice,* Blackie Academic & Professional, London.

L.L. McDonald, V.W. Lowe, R.K. Smidt, K.A. Meister (1976), A preliminary test for discriminant analysis based on small samples, *Biometrics* **32**, 417—422.

M. Moreira, E. Mayoraz (1998), Improved pairwise coupling classification with correcting classifiers. In *Proceedings of the 10th European Conference on Machine Learning (ECML-98)*, 160–171, Chemnitz, Germany, 1998.

M. Morisita (1959), Measuring of interspecific association and similarity between communities, *Mem. Fac. Sci. Kyushu Univ. Ser.* **E**, 65–80.

P.B. Nemenyi (1963), *Distribution-free multiple comparisons.* PhD thesis. Princeton University.

C.R. Rao (1965), *Linear Statistical Inference and Its Applications*, New York, Wiley.

A. Roy, R. Khattree (2005), On discrimination and classification with multivariate repeated measures data, *Journal of Statistical Planning and Inference* **134**, 462–485.

A. Roy, R. Khattree (2008), Classification rules for repeated measures data from biomedical research. In: *Computational methods in biomedical research, R. Khattree, D.N. Naik (Ed.)*, Chapman and Hall/CRC, 323–370.

B. Schölkopf, A. Smola, K.R. Müller (1998), Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation* **10**, 1299–1319.

R.H. Schumway, A.N. Unger (1974), Linear discriminant functions for stationary time series, *J. Amer. Statist. Assoc.* **69**, 948–956.

V. Vapnik (1998), *The nature of statistical learning theory*, New York, Springer Verlag.