

Wielowymiarowa funkcjonalna analiza regresji w zastosowaniu do klasyfikacji danych

Tomasz Górecki, Mirosław Krzyśko, Waldemar Wołyński

Wydział Matematyki i Informatyki
Uniwersytet im Adama Mickiewicza w Poznaniu

Statystyka Matematyczna
Będlewo 1-5.12.2014

Rozważmy klasyczny **problem klasyfikacyjny** polegający na wyznaczeniu procedury pozwalającej na przyporządkowanie danego obiektu do jednej z K populacji na podstawie obserwacji p cech charakteryzujących ten obiekt.

Klasyfikowany obiekt możemy opisać za pomocą pary losowej (\mathbf{X}, Y) , gdzie $\mathbf{X} = (X_1, X_2, \dots, X_p)' \in \mathbf{R}^p$ oraz $Y \in \{0, 1, \dots, K - 1\}$.

Zagadnienie klasyfikacyjne

W dalszym ciągu rozważmy jedynie przypadek $K = 2$. Istnieje wiele technik pozwalających na rozszerzenie zagadnienia binarnego na zagadnienie wieloklasowe (Krzyśko i Wołyński, 2009).

Wtedy

$$d(\mathbf{x}) = \begin{cases} 1, & P(Y = 1|\mathbf{X} = \mathbf{x}) \geq P(Y = 0|\mathbf{X} = \mathbf{x}); \\ 0, & P(Y = 1|\mathbf{X} = \mathbf{x}) < P(Y = 0|\mathbf{X} = \mathbf{x}). \end{cases}$$

Zauważmy, że

$$P(Y = 1|\mathbf{X} = \mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x}) = r(\mathbf{x}),$$

gdzie $r(\mathbf{x})$ jest **funkcją regresji** zmiennej losowej Y względem wektora losowego \mathbf{X} .

Stąd

$$d(\mathbf{x}) = \begin{cases} 1, & r(\mathbf{x}) \geq 1/2; \\ 0, & r(\mathbf{x}) < 1/2. \end{cases}$$

Założmy teraz, że klasyfikowany obiekt opisany jest **p -wymiarowym procesem losowym** $\mathbf{X}(t) = (X_1(t), X_2(t), \dots, X_p(t))'$, z ciągłym parametrem $t \in I$ oraz że $\mathbf{X}(t) \in L_2^p(I)$, gdzie $L_2(I)$ jest przestrzenią Hilberta funkcji całkownych z kwadratem.

Ponadto założmy, że proces ten podlega obserwacji w skończonej liczbie momentów czasowych oraz że proces transformacji od uzyskanych w ten sposób **danych dyskretnych** do **danych funkcjonalnych** wykonywany jest niezależnie dla każdej zmiennej.

Niech x_{kj} oznacza obserwowaną wartość zmiennej X_k , $k = 1, 2, \dots, p$ w j -tym momencie czasowym t_j , gdzie $j = 1, 2, \dots, J$. Zatem dane składają się pJ par postaci (t_j, x_{kj}) . Dane te podlegają procesowi wygładzania w wyniku którego otrzymujemy funkcje ciągłe $x_k(t)$, gdzie $t \in I$ oraz I oznacza przedział taki, że $t_j \in I$, gdzie $j = 1, \dots, J$ (Ramsay i Silverman, 2005).

Założmy, że funkcja $x_k(t)$ ma następującą reprezentację

$$x_k(t) = \sum_{b=0}^{B_k} c_{kb} \varphi_b(t), \quad t \in I, \quad k = 1, \dots, p,$$

gdzie $\{\varphi_b\}$ są ortonormalnymi funkcjami bazowymi, a c_{k0}, \dots, c_{kB_k} są nieznanymi współczynnikami.

Niech $\mathbf{x}(t) = (x_1(t), \dots, x_p(t))'$, $\mathbf{c} = (c_{10}, \dots, c_{1B_1}, \dots, c_{p0}, \dots, c_{pB_p})'$ oraz

$$\Phi(t) = \begin{bmatrix} \varphi'_1(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_2(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_p(t) \end{bmatrix},$$

gdzie $\varphi_k(t) = (\varphi_0(t), \dots, \varphi_{B_k}(t))'$, $k = 1, \dots, p$.

Wtedy

$$\mathbf{x}(t) = \Phi(t)\mathbf{c}.$$

Współczynniki c_{k0}, \dots, c_{kB_k} są estymowane metodą najmniejszych kwadratów, natomiast stopień gładkości funkcji $x_k(t)$ zależy od wartości B_k . Do wyznaczenia optymalnej wartości B_k możemy użyć bayesowskiego kryterium informacyjnego BIC (Shmueli, 2010).

Zajmijmy się teraz zagadnieniem estymacji funkcji regresji $r(\mathbf{x}(t))$.

Założmy, że dysponujemy *n*-elementową próbą uczącą

$$\mathcal{L}_n = \{(\mathbf{x}_1(t), y_1), (\mathbf{x}_2(t), y_2), \dots, (\mathbf{x}_n(t), y_n)\},$$

gdzie $\mathbf{x}_i(t) \in L_2^p(I)$ oraz $y_i \in \{0, 1\}$.

Przyjmujemy, że funkcje $\mathbf{x}_i(t)$ otrzymujemy jako wynik procesu wygładzania n niezależnych par dyskretnych danych (t_j, x_{kij}) , $k = 1, \dots, p$, $j = 1, \dots, J$, $i = 1, \dots, n$.

Zatem funkcje $\mathbf{x}_i(t)$ mają następujące reprezentacje

$$\mathbf{x}_i(t) = \Phi(t)\mathbf{c}_i, \quad i = 1, 2, \dots, n.$$

Model 1: Wielokrotna funkcjonalna regresja liniowa

Założmy następujący model funkcji regresji

$$r(\mathbf{x}(t)) = \beta_0 + \langle \boldsymbol{\beta}(t), \mathbf{x}(t) \rangle = \beta_0 + \int_I \boldsymbol{\beta}'(t) \mathbf{x}(t) dt.$$

Nieznanych parametrów funkcji regresji poszukujemy minimalizując sumę kwadratów

$$S(\beta_0, \boldsymbol{\beta}(t)) = \sum_{i=1}^n (y_i - \beta_0 - \int_I \boldsymbol{\beta}'(t) \mathbf{x}_i(t) dt)^2.$$

Przyjmujemy, że funkcje $\mathbf{x}_i(t)$, $i = 1, 2, \dots, n$ oraz $\boldsymbol{\beta}(t)$ leżą w tej samej przestrzeni oraz że

$$\mathbf{x}_i(t) = \boldsymbol{\Phi}(t) \mathbf{c}_i, \quad i = 1, 2, \dots, n, \quad \boldsymbol{\beta}(t) = \boldsymbol{\Phi}(t) \mathbf{d},$$

gdzie

$$\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_p(t))' \text{ oraz } \mathbf{d} = (d_{10}, \dots, d_{1B_1}, \dots, d_{p0}, \dots, d_{pB_p})'.$$

Ponieważ

$$\int_I \beta'(t) \mathbf{x}_i(t) dt = \mathbf{d}' \int_I \Phi'(t) \Phi(t) dt \mathbf{c}_i = \mathbf{d}' \mathbf{c}_i, \quad i = 1, 2, \dots, n,$$

to model wielokrotnej funkcjonalnej regresji liniowej możemy sprowadzić do standardowego modelu wielokrotnej regresji liniowej postaci

$$r(\mathbf{x}(t)) = \beta_0 + \mathbf{d}' \mathbf{c}.$$

Oznaczmy $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ oraz

$$\mathbf{Z} = \begin{bmatrix} 1 & \mathbf{c}'_1 \\ 1 & \mathbf{c}'_2 \\ \vdots & \vdots \\ 1 & \mathbf{c}'_n \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \beta_0 \\ \mathbf{d} \end{bmatrix}.$$

Jeżeli macierz $\mathbf{Z}'\mathbf{Z}$ jest nieosobliwa, to

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Model 2: Funkcjonalna regresja logistyczna

Przyjmujemy następujący model regresji dla danych funkcjonalnych

$$r(\mathbf{x}(t)) = \frac{\exp(\beta_0 + \langle \boldsymbol{\beta}(t), \mathbf{x}(t) \rangle)}{1 + \exp(\beta_0 + \langle \boldsymbol{\beta}(t), \mathbf{x}(t) \rangle)} = \frac{\exp(\beta_0 + \int_I \boldsymbol{\beta}'(t) \mathbf{x}(t) dt)}{1 + \exp(\beta_0 + \int_I \boldsymbol{\beta}'(t) \mathbf{x}(t) dt)}.$$

Postępując analogicznie jak w przypadku modelu wielokrotnej funkcjonalnej regresji liniowej, sprowadzamy powyższy model do standardowego modelu regresji logistycznej postaci

$$r(\mathbf{x}(t)) = \frac{\exp(\beta_0 + \mathbf{d}'\mathbf{c})}{1 + \exp(\beta_0 + \mathbf{d}'\mathbf{c})}.$$

Po estymacji nieznanymi parametrów, otrzymujemy następującą postać estymatora funkcji regresji w modelu funkcjonalnej regresji logistycznej

$$\hat{r}(\mathbf{x}(t)) = \frac{\exp(\hat{\beta}_0 + \hat{\mathbf{d}}'\mathbf{c})}{1 + \exp(\hat{\beta}_0 + \hat{\mathbf{d}}'\mathbf{c})}.$$

Model 3: Wygładzarka liniowa dla wielowymiarowych danych funkcjonalnych

Niech $\mathbf{x}_0(t)$ będzie ustalonym i znanym punktem w przestrzeni $L_2^p(I)$.
Rozwijamy funkcję $r(\mathbf{x}(t))$ w szereg Taylora w otoczeniu punktu $\mathbf{x}_0(t)$.
Mamy

$$\begin{aligned} r(\mathbf{x}(t)) &\approx r(\mathbf{x}_0(t)) + \left\langle \frac{\partial r(\mathbf{x}_0(t))}{\partial \mathbf{x}_0(t)}, \mathbf{x}(t) - \mathbf{x}_0(t) \right\rangle \\ &= a + \langle \mathbf{b}(t), \mathbf{x}(t) - \mathbf{x}_0(t) \rangle, \end{aligned}$$

gdzie

$$a = r(\mathbf{x}_0(t)), \quad \mathbf{b}(t) = \frac{\partial r(\mathbf{x}_0(t))}{\partial \mathbf{x}_0(t)}.$$

Model 3: Wygładzarka liniowa dla wielowymiarowych danych funkcjonalnych

Nieznanych parametrów funkcji regresji poszukujemy minimalizując ważoną sumę kwadratów postaci

$$S(a, \mathbf{b}(t)) = \sum_{i=1}^n \left(y_i - a - \int_I \mathbf{b}'(t)(\mathbf{x}_i(t) - \mathbf{x}_0(t)) dt \right)^2 K \left(\frac{1}{h} \|\mathbf{x}_i(t) - \mathbf{x}_0(t)\| \right)$$

gdzie K jest funkcją jądrową, a $h > 0$ parametrem wygładzającym.

Najczęściej wybiera się jądro Epanechnikowa

$$K(z) = \frac{3}{4} (1 - z^2) I_{\{|z| \leq 1\}}$$

lub jądro gaussowskie

$$K(z) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}z^2\right), \quad z \in \mathbb{R}.$$

Model 3: Wygładzarka liniowa dla wielowymiarowych danych funkcjonalnych

W dalszym ciągu zakładamy, że funkcje $\mathbf{x}_i(t)$, $i = 1, 2, \dots, n$ oraz $\mathbf{b}(t)$ leżą w tej samej przestrzeni oraz że

$$\mathbf{x}_i(t) = \Phi(t)\mathbf{c}_i, \quad i = 1, 2, \dots, n, \quad \mathbf{b}(t) = \Phi(t)\mathbf{d}.$$

Oznaczmy $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ oraz

$$\mathbf{Z} = \begin{bmatrix} 1 & (\mathbf{c}_1 - \mathbf{c}_0)' \\ 1 & (\mathbf{c}_2 - \mathbf{c}_0)' \\ \vdots & \vdots \\ 1 & (\mathbf{c}_n - \mathbf{c}_0)' \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} a \\ \mathbf{d} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} K_1 & 0 & \dots & 0 \\ 0 & K_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_n \end{bmatrix}.$$

Model 3: Wygładzarka liniowa dla wielowymiarowych danych funkcjonalnych

Jeżeli macierz $Z'WZ$ jest nieosobliwa, to

$$\hat{\gamma} = (Z'WZ)^{-1}Z'Wy.$$

Model 3: Wygładzarka liniowa dla wielowymiarowych danych funkcjonalnych

W rozwinięciu w szereg Taylora funkcji $r(\mathbf{x}(t))$ w otoczeniu punktu $\mathbf{x}_0(t)$ weźmy pod uwagę jedynie pierwszy wyraz, tzn.

$$r(\mathbf{x}(t)) \approx r(\mathbf{x}_0(t)) = a.$$

Nieznaną wartość parametru a wyznaczamy poprzez minimalizację ważonej sumy kwadratów postaci

$$S(a) = \sum_{i=1}^n (y_i - a)^2 K\left(\frac{1}{h} \|\mathbf{x}_i(t) - \mathbf{x}_0(t)\|\right).$$

W rezultacie otrzymujemy

$$\hat{r}(\mathbf{x}(t)) = \frac{\sum_{i=1}^n y_i K_i}{\sum_{i=1}^n K_i},$$

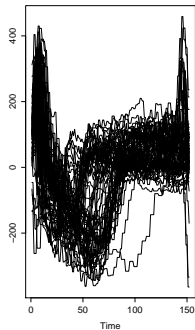
gdzie $K_i = K\left(\frac{1}{h} \sqrt{(\mathbf{c}_i - \mathbf{c}_0)'(\mathbf{c}_i - \mathbf{c}_0)}\right)$, $i = 1, 2, \dots, n$.

Jest to znany **estymator jądrowy Nadarayi-Watsona**.

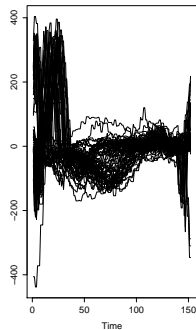
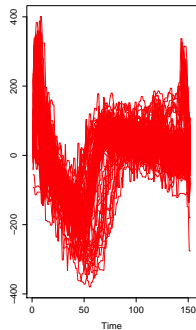
Dane są zapisem EKG z dwóch elektrod i dotyczą 200 pacjentów podzielonych na 2 grupy: 133 pacjentów zdrowych oraz 67 pacjentów z nadkomorowym przedwczesnym rytmem serca.

Długość szeregów czasowych wynosiła 152 punktów czasowych.

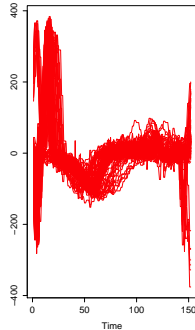
Przykład 1: EKG



Cecha 1



Cecha 2

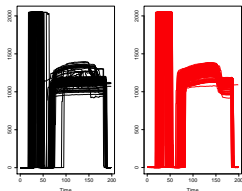


Przykład 2: *Wafle krzemowe*

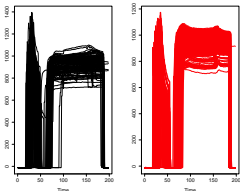
Dane pochodzą z zakładu produkcji półprzewodników i dotyczą produkcji wafli krzemowych. 6 czujników zainstalowanych w komorach próżniowych rejestrowało różne parametry procesu produkcyjnego.

Zarejestrowano 327 szeregów czasowych podzielonych na dwie grupy: 200 szeregów odpowiadających wafłom prawidłowym i 127 szeregów odpowiadających wafłom wadliwym. Długość szeregów wynosiła 198 punktów czasowych.

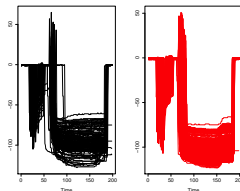
Przykład 2: Wafle krzemowe



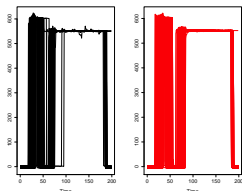
Cecha 1



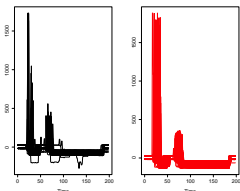
Cecha 2



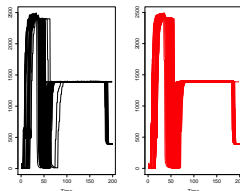
Cecha 3



Cecha 4









Cecha 5



Cecha 6

Dla każdego ze zbiorów policzony został błąd klasyfikacji metodą sprawdzania krzyżowego (LOO CV).

Metoda	EKG	Wafle krzemowe
wielokrotna funkcjonalna regresja liniowa	11.50	0.59
funkcjonalna regresja logistyczna	11.50	0.17
wygładzarka liniowa dla danych funkcjonalnych	16.50	0.67
estymator jądrowy Nadarayi-Watsona	20.50	10.64

-  Anderson, T.W. (1984): An Introduction to Multivariate Statistical Analysis. Wiley.
-  Horvath, L., Kokoszka, P. (2012): Inference for Functional Data with Applications, Springer, New York.
-  Krzyśko, M., Wołyński, W. (2009): New variants of pairwise classification. European Journal of Operational Research 199(2), 512-519.
-  Muller, H.G., Stadtmuller, U. (2005): Generalized functional linear models. Annals of Statistics 33, 774805.
-  Ramsay, J.O., Silverman, B.W. (2005): Functional Data Analysis, Second Edition, Springer.
-  Shmueli, G. (2010): To explain or to predict?, Statistical Science 25(3), 289-310.