J. PERKAL (Wrocław)

O DYSKRYMINACJI PŁASKIEJ⁽¹⁾

1. Wstęp.

1.1. Podczas współpracy matematyczno-przyrodniczej często natrafiamy na pewien charakterystyczny konflikt. Przyrodnik, chcąc poznać przebieg lub stan jakiegoś zjawiska, zbiera odpowiednie dane pomiarowe, przychodzi z nimi do matematyka i prosi o pomoc w opracowaniu i w wyciąganiu wniosków. Matematyka zagadnienie zaciekaawi dopiero wtedy, gdy dojrzy w nim sens matematyczny, prosi więc przyrodnika o zadawanie takich pytań, na które mógłby dać optymalną odpowiedź. Tak np. przyrodnik, pomierzywszy zbiór indywiduów, chce, aby matematyk pomógł mu ocenić podobieństwa i różnice między indywiduami, uporządkować te indywidua i pogrupować w klasy. Matematyka nie interesuje tak ogólnie postawione zagadnienie. Ale część tego zadania, mianowicie podział na klasy, jest znanym, optymalnie rozwiązany przez R. A. Fishera zagadnieniem dyskryminacji. Trzeba tylko założyć, że funkcją dyskryminującą może być jedynie funkcja liniowa. Przyrodnik, często nie rozumiejąc przyrodniczego znaczenia tego założenia, godzi się na nie, przyrzekając sobie, że wyniki, jakie otrzyma od matematyka, potraktuje jako częściowe rozwiązanie swego zagadnienia. Ale matematyk odsuwa przyrodnika od dalszych trudnych rachunków i dostarcza mu gotowe wyniki, tj. wykaz numerów indywiduów w poszczególnych grupach. Przyrodnik czuje się rozczarowany. Jego obszerne, bogate doświadczenie, mające wzbogacić wiedzę przyrodniczą, zostało ograniczone i zutilitaryzowane. Metoda, którą zastosował matematyk, jest optymalna dla pewnych praktycznych celów. Założenia niezamierzone przez przyrodnika, lecz tylko przyjęte na skutek trudności matematycznych, spowodowały obiektywizację, która nie dopuszcza żadnych pretensji i wątpliwości przyrodnika, gdy ten chciałby np. zamienić któreś indywiduum jednej grupy na któreś innej grupy. Jeśli przyrodnik zgodził się na założenia,

⁽¹⁾ Z materiałów pośmiertnych do druku przygotował J. Łukasiewicz.

których listę matematyk mu przedstawił, to teraz nie ma miejsca na dyskusje. Wyniki matematyka są jednoznaczne. Jeśli przyrodnik kwestionuje wyniki, to znaczy, że nie zgadza się na któreś z założeń. A matematyk nie chce współpracować z przyrodnikiem, który w trakcie rachunków wycofuje się z przyjętych uprzednio założeń. Oczywiście, formalnie rzecz biorąc, matematyk ma rację, ale tu przecież nie o rację idzie, lecz o to, żeby pomóc przyrodnikowi, a jemu wcale nie zależało na założeniach proponowanych przez matematyka. Przyjął je jako zło konieczne, jako pewną aproksymację rzeczywistości, bez której matematyk ze swym aparatem myślowym i maszyną rachunkową był bezsilny. Przyrodnik jednak chce za najwyższy argument uważać swój zdrowy rozsądek i zgodność wyników ze swoimi koncepcjami.

W rezultacie przyrodnicy rzadko są zadowoleni z tego rodzaju współpracy i rzadko stosują klasyczną metodę dyskryminacji do podziału zbioru indywiduów na więcej niż dwie grupy. Częściej szukają innych rozwiązań mniej lub bardziej poprawnych, lecz nie tak mechanicznych (patrz np. [1] i [4]).

1.2. Podsunąłem już przyrodnikom kilka pomysłów mniej mechanicznej dyskryminacji (patrz [2] i [3]), jak np. dyskryminację uproszczoną i dyskryminację dendrytową. Obecnie proponuję jeszcze dwie metody dyskryminacji płaskiej. Polegają one zasadniczo na tym, że badane zbiory punktów przenosimy z n -wymiarowej przestrzeni euklidesowej (n jest równe liczbie mierzonych cech) na płaszczyznę, i to w ten sposób, żeby dały się możliwie dobrze rozdzielić. Zyskujemy w ten sposób pogładowość i łatwość interpretacji, gdyż przyrodnik ma cały swój materiał na jednym rysunku. Co prawda rezygnujemy przy takiej klasyfikacji z części informacji zawartej w materiale pomiarów. Poprzestajemy jedynie na tej, która zawiera się na rysunku, czyli w dwóch współrzędnych (parametrach). Jest to jednak duża część informacji, a pozwalając sobie na niewielką demagogię, zauważamy, że funkcja Fishera ogranicza się do informacji zawartej w jednym tylko parametrze.

1.3. Fisher, wprowadzając swoją funkcję dyskryminacyjną $f(z)$, żąda, aby dla dwóch zbiorów Z_1 i Z_2 o średnich z_1 i z_2 funkcja ta przyjmowała odległe wartości, czyli aby $|f(z_1) - f(z_2)|$ było duże. My w przypadku K zbiorów $Z_1, Z_2, \dots, Z_K$ o średnich z_k (gdzie $k = 1, 2, \dots, K$; $z_k = (z_{1k}, z_{2k}, \dots, z_{Ik})$; I jest liczbą współrzędnych, czyli cech charakteryzujących każdy punkt) będziemy żądali, aby duża była suma kwadratów różnic między wielkościami $f(z_k)$

$$S^2 = \sum_{k=1}^K \sum_{h=1}^K [f(z_k) - f(z_h)]^2.$$

Oznaczając przez $\bar{f} = \frac{1}{K} \sum_{k=1}^K f(z_k)$ średnią arytmetyczną wielkości $f(z_k)$, otrzymamy

$$\begin{aligned} S^2 &= \sum_{k=1}^K \sum_{h=1}^K [f(z_k) - \bar{f} + \bar{f} - f(z_h)]^2 = \\ &= \sum_{k=1}^K \sum_{h=1}^K \{[f(z_k) - \bar{f}]^2 + [f(z_h) - \bar{f}]^2 - 2[f(z_k) - \bar{f}][f(z_h) - \bar{f}]\}. \end{aligned}$$

Oznaczając następnie przez $S_f^2 = \frac{1}{K} \sum_{k=1}^K [f(z_k) - \bar{f}]^2$ wariancję wartości $f(z_k)$ i uwzględniając, że

$$\sum_{k=1}^K \sum_{h=1}^K [f(z_k) - \bar{f}][f(z_h) - \bar{f}] = \sum_{k=1}^K [f(z_k) - \bar{f}] \cdot \sum_{h=1}^K [f(z_h) - \bar{f}] = 0,$$

otrzymamy ostatecznie

$$S^2 = 2K^2 S_f^2.$$

Jak widać, postulat maksymizacji S^2 jest równoważny warunkowi, aby wariancja wartości $f(z_k)$ była jak największa. Tego będziemy żądać, mimo że warunek ten nie musi prowadzić do zróżnicowania średnich w poszczególnych zbiorach. Tak więc funkcja dyskryminacyjna Fishera musiała rozsuwać średnie rozdzielanych dwóch zbiorów, gdy tymczasem nasze metody dyskryminacyjne nie mówią a priori o tym, że rozdzielane zbiory będą miały różne wartości funkcji dyskryminującej dla punktów średnich w zbiorach. Mogłoby się zdarzyć (przy dyskryminacji co najmniej trzech zbiorów), że na rysunku dyskryminującym w tym samym punkcie znalazłyby się średnie dwóch zbiorów. Tylko przyrodoznik mógłby wtedy stwierdzić, czy dyskryminacja się udała.

1.4. Weźmy pod uwagę próbki $Z_k (k = 1, 2, \dots, K)$, z których każda zawiera J osobników o numerach $j = 1, 2, \dots, J$. Każdy osobnik jest określony I cechami o numerach $i = 1, 2, \dots, I$. Wielkość i -tej cechy j -tego osobnika k -tej próbki (grupy) oznaczmy przez z_{ijk} . Zastąpienie wskaźnika (i, j lub k) kropką będzie oznaczało średniowanie; tak np.

$z_{i,k} = \frac{1}{J} \sum_{j=1}^J z_{ijk}$. Podobnie określamy $z_{i..}$, $z_{..k}$ lub $z_{...}$. Oznaczmy przez $\mathbf{Z}$

macierz o I kolumnach i $J \cdot K$ wierszach, gdzie w i -tej kolumnie i $[j + (k-1)J]$ -tym wierszu znajduje się element z_{ijk} . Przez $\bar{\mathbf{Z}}$ oznaczmy macierz tego samego formatu o elementach $z_{i,k}$ w i -tej kolumnie i w wierszach o numerach od $(k-1)J+1$ do kJ . Wreszcie przez $\bar{\bar{\mathbf{Z}}}$ oznaczmy

macierz także tego samego formatu, w której wszystkie elementy i -tej kolumny są równe $z_{i..}$. Macierze drugich momentów

$$W = \frac{1}{KJ} (\mathbf{Z} - \bar{\mathbf{Z}})' (\mathbf{Z} - \bar{\mathbf{Z}}); \quad \mathbf{M} = \frac{1}{KJ} (\bar{\mathbf{Z}} - \bar{\bar{\mathbf{Z}}})' (\bar{\mathbf{Z}} - \bar{\bar{\mathbf{Z}}})$$

będziemy nazywali odpowiednio macierzą wewnątrzgrupową i macierzą międzygrupową. Elementami tych macierzy będą

$$w_{ij} = \frac{1}{KJ} \sum_{k=1}^K \sum_{l=1}^J (z_{ilk} - z_{i,k})(z_{jlk} - z_{j,k});$$

$$m_{ij} = \frac{1}{K} \sum_{k=1}^K (z_{i,k} - z_{i..})(z_{j,k} - z_{j..}).$$

Symbolem $\mathbf{z} = (z_1, z_2, \dots, z_I)$ oznaczmy wektor o I składowych, a symbolem z_{jk} odpowiedni wektor (wiersz o numerze $(k-1)J+j$) macierzy $\mathbf{Z}$. Średnie wektory $z_{.k}$ oraz $z_{..}$ są odpowiednimi wierszami macierzy $\bar{\mathbf{Z}}$ i $\bar{\bar{\mathbf{Z}}}$.

2. Metoda optymalna.

2.1. Znajdziemy teraz dwa prostopadłe wektory jednostkowe

$$\mathbf{a} = (a_1, a_2, \dots, a_I); \quad \mathbf{b} = (b_1, b_2, \dots, b_I),$$

czyli wektory spełniające warunki

$$(1) \quad \mathbf{a}\mathbf{b}' = 0 \quad \text{oraz} \quad \mathbf{a}\mathbf{a}' = \mathbf{b}\mathbf{b}' = 1,$$

od których zażądamy, żeby kombinacje liniowe

$$x = \mathbf{a}\mathbf{z}' \quad \text{oraz} \quad y = \mathbf{b}\mathbf{z}'$$

dawały dobrze dyskryminującą płaszczyznę.

Otrzymamy w ten sposób przekształcenie, które obserwacjom wielocechowym, czyli wektorom $\mathbf{z}$, będzie przyporządkowywało parę liczb x i y , czyli punkt płaszczyzny (x, y) .

$$(x, y) = \mathbf{f}(\mathbf{z}) = (\mathbf{a}\mathbf{z}', \mathbf{b}\mathbf{z}').$$

Metoda będzie dobrze rozdzielała grupy (próbki), jeśli małe będą odległości między punktami (x, y) przedstawiającymi osobniki tej samej grupy, a duże będą odległości między punktami przedstawiającymi średnie różnych grup. Oznaczmy

$$(2) \quad d = \frac{1}{KJ} \sum_{k=1}^K \sum_{j=1}^J [(x_{jk} - x_{.k})^2 + (y_{jk} - y_{.k})^2],$$

gdzie

$$(x_{jk}, y_{jk}) = \mathbf{f}(\mathbf{z}_{jk}) \quad \text{i} \quad (x_{.k}, y_{.k}) = \mathbf{f}(\mathbf{z}_{.k}),$$

oraz

$$(3) \quad n = \frac{1}{K} \sum_{k=1}^K [(x_{.k} - x_{..})^2 + (y_{.k} - y_{..})^2],$$

gdzie

$$(x_{..}, y_{..}) = \mathbf{f}(\mathbf{z}_{..}).$$

Jako warunek optymalności dyskryminacji przyjmujemy

$$(4) \quad p = \frac{n}{d} = \text{maksimum}.$$

Wektory $\mathbf{a}$ i $\mathbf{b}$ dobierzemy w ten sposób, aby przy spełnieniu warunków (1) realizowały postulat (4).

2.2. Wzór (2) można przekształcić w następujący sposób:

$$(5) \quad d = \frac{1}{KJ} \sum_{k=1}^K \sum_{j=1}^J \left\{ \left[\sum_{i=1}^I a_i (z_{ijk} - z_{i,k}) \right]^2 + \left[\sum_{i=1}^I b_i (z_{ijk} - z_{i,k}) \right]^2 \right\} = \\ = \sum_{i=1}^I \sum_{j=1}^J (a_i a_j w_{ij} + b_i b_j w_{ij}) = \mathbf{a} \mathbf{W} \mathbf{a}' + \mathbf{b} \mathbf{W} \mathbf{b}'.$$

Podobne przekształcenie doprowadzi (3) do postaci

$$(6) \quad n = \mathbf{a} \mathbf{M} \mathbf{a}' + \mathbf{b} \mathbf{M} \mathbf{b}'.$$

Wobec tego warunek (4) polega na znalezieniu ekstremum wyrażenia

$$(7) \quad p = \frac{n}{d} = \frac{\mathbf{a} \mathbf{M} \mathbf{a}' + \mathbf{b} \mathbf{M} \mathbf{b}'}{\mathbf{a} \mathbf{W} \mathbf{a}' + \mathbf{b} \mathbf{W} \mathbf{b}'}.$$

2.3. Ekstremum wyrażenia (4) przy warunkach (1) znajdziemy metodą mnożników Lagrange'a λ'_1, λ'_2 i λ'_3 . Tworzymy funkcję Lagrange'a

$$F(\mathbf{a}, \mathbf{b}) = \frac{n}{d} - \lambda'_1 \mathbf{a} \mathbf{a}' - \lambda'_2 \mathbf{b} \mathbf{b}' - \lambda'_3 \mathbf{a} \mathbf{b}'$$

i pochodne cząstkowe tej funkcji względem a_i oraz b_i przyrównamy do zera

$$\frac{\partial F}{\partial a_i} = \frac{1}{d^2} \left(d \frac{\partial n}{\partial a_i} - n \frac{\partial d}{\partial a_i} \right) - 2\lambda'_1 a_i - \lambda'_3 b_i = 0, \\ \frac{\partial F}{\partial b_i} = \frac{1}{d^2} \left(d \frac{\partial n}{\partial b_i} - n \frac{\partial d}{\partial b_i} \right) - 2\lambda'_2 b_i - \lambda'_3 a_i = 0.$$

Mnożąc te równania przed d i podstawiając

$$2\lambda_1 = 2\lambda'_1 d, \quad 2\lambda_2 = 2\lambda'_2 d, \quad 2\lambda_3 = \lambda'_3 d,$$

otrzymamy

$$\begin{aligned} \frac{\partial n}{\partial a_i} - p \frac{\partial d}{\partial a_i} &= 2\lambda_1 a_i + 2\lambda_3 b_i, \\ \frac{\partial n}{\partial b_i} - p \frac{\partial d}{\partial b_i} &= 2\lambda_2 b_i + 2\lambda_3 a_i. \end{aligned}$$

Jeśli przez

$$\mathbf{n}_a = \left(\frac{\partial n}{\partial a_1}, \frac{\partial n}{\partial a_2}, \dots, \frac{\partial n}{\partial a_I} \right)$$

oznaczymy wektor złożony z I pochodnych cząstkowych wielkości n względem a_i oraz analogicznie wprowadzimy oznaczenia $\mathbf{n}_b$, $\mathbf{d}_a$ i $\mathbf{d}_b$, to ostatni układ równań zapiszemy krócej w postaci

$$\begin{aligned} (8) \quad \mathbf{n}_a - p \mathbf{d}_a &= 2\lambda_1 \mathbf{a} + 2\lambda_3 \mathbf{b}, \\ \mathbf{n}_b - p \mathbf{d}_b &= 2\lambda_2 \mathbf{b} + 2\lambda_3 \mathbf{a}. \end{aligned}$$

2.4. Z równań (5) i (6) lub, jeśli kto woli, różniczkując wyrażenia (2) i (3) względem a_i i względem b_i , otrzymamy

$$\begin{aligned} (9) \quad \mathbf{n}_a &= 2\mathbf{a}M, \quad \mathbf{d}_a = 2\mathbf{a}W, \\ \mathbf{n}_b &= 2\mathbf{b}M, \quad \mathbf{d}_b = 2\mathbf{b}W. \end{aligned}$$

Podstawiając te wartości do równań (8), dostaniemy

$$\mathbf{a}P = \lambda_1 \mathbf{a} + \lambda_3 \mathbf{b}, \quad \mathbf{b}P = \lambda_2 \mathbf{b} + \lambda_3 \mathbf{a},$$

gdzie

$$(10) \quad P = M - pW$$

jest symetryczną macierzą kwadratową formatu $I \times I$ (bo M i W są macierzami symetrycznymi).

Gdybyśmy zrezygnowali z warunku ortogonalności wektorów $\mathbf{a}$ i $\mathbf{b}$, to moglibyśmy przyjąć $\lambda_3 = 0$, a wtedy układ ten sprowadziłby się do równań

$$\mathbf{a}P = \lambda_1 \mathbf{a}, \quad \mathbf{b}P = \lambda_2 \mathbf{b},$$

z których wynika, że $\mathbf{a}$ i $\mathbf{b}$ byłyby wektorami własnymi macierzy P odpowiadającymi wartościom własnym λ_1 i λ_2 . Jako takie, wektory $\mathbf{a}$ i $\mathbf{b}$ byłyby ortogonalne.

Jeśli wartości własne odpowiadające wektorom $\mathbf{a}$ i $\mathbf{b}$ oznaczmy przez λ_a i λ_b , to z (7) i (10) dostaniemy

$$\mathbf{aPa}' + \mathbf{bPb}' = 0,$$

czyli że odpowiednie wartości własne muszą spełniać warunek

$$(11) \quad \lambda_a + \lambda_b = 0.$$

Rozwiązaniem będą więc dwa różne wektory własne $\mathbf{a} \neq \mathbf{b}$ macierzy $\mathbf{P}$, odpowiadające wartościom własnym λ_a , λ_b spełniającym warunek (11).

Wielkość $p = n/d$ jest stosunkiem rozsiewu środków grup do rozsiewu punktów wewnątrz grup. Nazwiemy tę wielkość *siłą dyskryminacji*. Jeśli siła dyskryminacji jest duża (np. $p > 3$), to należy oczekiwać, że rysunek dyskryminacyjny będzie dobrze rozdzielał grupy. Jeśli siła dyskryminacji jest mała (np. $p < 1$), to nie należy oczekiwać efektu rozdzielania grup. Byłoby ciekawe zbadać rozkład siły dyskryminacyjnej p w przypadku K grup wylosowanych z łącznego rozkładu normalnego.

2.5. Rozwiązanie sprowadza się więc do następującego rachunku: Znając macierze $\mathbf{M}$ i $\mathbf{W}$ należy tak dobrać największą liczbę p , żeby macierz $\mathbf{P} = \mathbf{M} - p\mathbf{W}$ miała wektory własne $\mathbf{a}$ i $\mathbf{b}$ odpowiadające wartościom własnym λ_a , λ_b spełniającym warunek (11). Wektory $\mathbf{a}$, $\mathbf{b}$ spełniają wtedy automatycznie warunek (4).

3. Metoda dwuzespołowa.

3.1. Metodę tę należy stosować, gdy rozważane indywidua scharakteryzowane są dwoma zespołami cech: $\mathbf{z}_1 = (z_1, z_2, \dots, z_{I_1})$ oraz $\mathbf{z}_2 = (z_{I_1+1}, z_{I_1+2}, \dots, z_{I_1+I_2})$, przy czym $I_1 + I_2 = I$. Oczywiście każdy zespół I cech można zawsze rozważać jako układ dwóch takich zespołów. Niekiedy jednak przyrodnik może uważać, że cechy dzielą się w naturalny sposób na dwa zespoły. Niekiedy do wniosku takiego dojdzie matematyk po analizie taksonomicznej (np. z pomocą dendrytów dualnych) lub po analizie macierzy drugich momentów (zespoły zgodne).

Macierze $\mathbf{Z}$, $\mathbf{Z}$, $\mathbf{Z}$, jak również macierze drugich momentów $\mathbf{W}$ i $\mathbf{M}$ będą obecnie zaopatrzone wskaźnikami 1 lub 2 zależnie od tego, czy dotyczą pierwszego, czy drugiego zespołu cech.

3.2. Jako współrzędnych punktów na płaszczyźnie dyskryminacyjnej użyjemy teraz dwóch wielkości x i y , z których pierwszą obliczamy jako funkcję liniową pierwszego zespołu cech najlepiej dyskryminującą dane K próbek, a drugą obliczamy podobnie z drugiego zespołu cech. Jest więc

$$\begin{aligned} x &= \mathbf{az}'_1, & \text{gdzie} & \quad \mathbf{a} = (a_1, a_2, \dots, a_{I_1}), \\ y &= \mathbf{bz}'_2, & \text{gdzie} & \quad \mathbf{b} = (b_1, b_2, \dots, b_{I_2}). \end{aligned}$$

Wektor $\mathbf{a}$ dobieramy w ten sposób, aby było $\mathbf{a}\mathbf{a}' = 1$ oraz aby osiągało maksimum wyrażenie

$$p_1 = \frac{n_1}{d_1},$$

gdzie

$$d_1 = \frac{1}{KJ} \sum_{k=1}^K \sum_{j=1}^J (x_{jk} - x_{.k})^2 = \mathbf{a}\mathbf{W}_1\mathbf{a}',$$

$$n_1 = \frac{1}{K} \sum_{k=1}^K (x_{.k} - x_{..})^2 = \mathbf{a}\mathbf{M}_1\mathbf{a}'.$$

Funkcją Lagrange'a będzie tu

$$F_1(\mathbf{a}) = \frac{n_1}{d_1} - \lambda' \mathbf{a}\mathbf{a}',$$

a przyrównanie do zera jej pochodnej pomnożonej przez d_1 daje

$$\frac{\partial n_1}{\partial a_i} - p_1 \frac{\partial d_1}{\partial a_i} = 2\lambda a_i, \quad \text{gdzie} \quad \lambda = \lambda' d_1.$$

Po przekształceniu równanie to możemy napisać w postaci

$$\sum_{j=1}^{I_1} a_j m_{1ij} - p_1 \sum_{j=1}^{I_1} a_j w_{1ij} = \lambda a_i$$

albo w postaci macierzowej

$$\mathbf{a}\mathbf{M}_1 - p_1 \mathbf{a}\mathbf{W}_1 = \lambda \mathbf{a},$$

lub ostatecznie

$$(10) \quad \mathbf{a}\mathbf{P}_1 = \lambda \mathbf{a},$$

gdzie $\mathbf{P}_1 = \mathbf{M}_1 - p_1 \mathbf{W}_1$.

Okazuje się więc, że wektor $\mathbf{a}$ jest wektorem własnym macierzy $\mathbf{P}_1$, a dla maksymalnego

$$p_1 = \frac{\mathbf{a}\mathbf{M}_1\mathbf{a}'}{\mathbf{a}\mathbf{W}_1\mathbf{a}'}$$

z równania (10) wynika $\mathbf{a}\mathbf{P}_1\mathbf{a}' = 0$, czyli $\lambda = 0$.

Zadanie znalezienia wektora $\mathbf{a}$ rozwiązuje się podobnie jak w punkcie 2.5. Znając $\mathbf{M}_1$ i $\mathbf{W}_1$ dobieramy p_1 tak, żeby wektor własny $\mathbf{a}$ macierzy $\mathbf{P}$ spełniał równanie (10) lub żeby wartość własna λ tej macierzy była równa zeru.

Zupełnie analogicznie wyznaczamy wektor własny macierzy $P_2 = M_2 - p_2 W_2$ spełniający warunek

$$p_2 = \frac{\mathbf{b}M_2\mathbf{b}'}{\mathbf{b}W_2\mathbf{b}'} = \text{maksimum}.$$

3.3. Wielkości p_1 i p_2 można nazwać siłami dyskryminacyjnymi współrzędnych x i y płaszczyzny dyskryminacyjnej. Łączną siłą dyskryminacyjną metody dwuzespołowej będzie

$$(11) \quad p = \frac{\mathbf{a}M_1\mathbf{a}' + \mathbf{b}M_2\mathbf{b}'}{\mathbf{a}W_1\mathbf{a}' + \mathbf{b}W_2\mathbf{b}'}.$$

3.4. Zarówno dyskryminacja optymalna, jak dwuzespołowa mają dwie wady. Pierwszą wadą jest to, że oba rysunki dyskryminacyjne są pewnymi nie naturalnymi, lecz tendencyjnymi obrazami zbioru punktów indywidualnych I -wymiarowej przestrzeni. Obrazy te od razu konstruujemy tak, żeby na nich poszczególne grupy punktów dały się rozdzielić. Oczywiście to było naszym celem, ale pozostaje pewna wątpliwość, czy do osiągnięcia tego celu nie wykorzystaliśmy tzw. błędu próbki, przypadkowych właściwości posiadanego akurat materiału. Tę wadę ma również metoda dyskryminacji Fishera, na co zwracaliśmy uwagę na początku. Metoda Fishera jest nawet bardziej narażona na błąd próbki, gdyż wyzyskuje wszystkie różnice między średnimi w grupach, gdy tymczasem opisane tu metody wyzyskują zamiast tego tylko momenty międzygrupowe. Drugą wadą są trudności rachunkowe. Wielkości p są określone za pomocą wektorów $\mathbf{a}$ i $\mathbf{b}$, będących wektorami własnymi macierzy zdefiniowanej tą samą wielkością p . Wymaga to stosowania metody kolejnych przybliżeń za pomocą elektronicznej maszyny cyfrowej i dość skomplikowanego programu.

Prostą i naturalną metodą dyskryminacji dwuzespołowej nie mającą tych wad, lecz za to na ogół gorzej rozdzielającą grupy, jest metoda dwuwskaznikowa, w której oznaczamy przez $x = m_1$ i przez $y = m_2$ wskaźniki główne (patrz [3]) odpowiednio pierwszego i drugiego zespołu cech.

Prace cytowane

[1] T. Bogdanik, *Zastosowanie dyskryminacji matematycznej do różnicowania dwóch typów cukrzycy*, Polski Tygodnik Lekarski 15 (1960), nr 31.

[2] J. Perkal, *Two new methods of analysing a collection of attributes*, Bulletin de l'Académie Polonaise de Sciences 7 (1959), str. 63-66.

[3] —, *On the analysis of a set of characteristics*, Zastosow. Matem. 5 (1960), str. 35-45.

[4] —, *Matematyka dla przyrodników i rolników*, tomy II i III, PWN, Warszawa 1963.

Praca wpłynęła 16. 9. 1965

Ю. ПЭРКАЛЬ (Вроцлав)

О ПЛОСКОЙ ДИСКРИМИНАЦИИ

РЕЗЮМЕ

Автор предлагает два новых метода дискриминации биологических индивидов. В каждом из них совокупность I признаков ($I > 2$), описывающих классифицированные индивиды, сводится к двум параметрам, допускающим наглядное представление индивидов в виде точек плоскости. При этом желается достигнуть хорошее разделение данных выборок Z_k ($k = 1, 2, \dots, K$), состоящих каждая из одного и того же самого числа индивидов.

Так называемый *оптимальный метод* требует вычисления линейных преобразований $x = a \cdot z'$ и $y = b \cdot z'$, где $z = (z_1, z_2, \dots, z_I)$ обозначает вектор представляющий исследованного индивида в I -мерном пространстве количественных признаков, а векторы a и b удовлетворяют условиям (1) и (4). Второе из них дает максимальное значение частного от дисперсии (3) преобразованных средних отдельных выборок и внутренней дисперсии (2) во всех преобразованных выборках.

Второе предложение — это *два-коллективный метод*, состоящий в том, что совокупность I признаков разделяется на два коллектива. В каждом из них независимо вычисляется линейная дискриминационная функция.

J. PERKAL (Wrocław)

ON TWO-DIMENSIONAL DISCRIMINATION METHODS

SUMMARY

The author proposes two new methods for the discrimination of individuals. In both of them the set of I ($I > 2$) characteristics, used for the description of classified individuals, is reduced to two parameters which enable the graphical presentation of data in a two-dimensional coordinate system. It has to be done in such a way as to obtain a good separation of given samples Z_k ($k = 1, 2, \dots, K$), each of them containing the same number of individuals.

The so called *optimum method* consists in the determination of linear transformations $x = a \cdot z'$ and $y = b \cdot z'$, where $z = (z_1, z_2, \dots, z_I)$ is the vector representing a classified individual and vectors a, b satisfy conditions (1) and (4). Condition (7) requires the maximum value of the quotient of the variance (3) of transformed sample means and the inner variance (2) inside transformed samples.

The second proposal is the *two-collective method* consisting in a partition of the set of I characteristics into two collectives. Within each of them a linear discrimination function is then independently determined.