ISSN 1507-3866

Paweł Siarka
Institute of Financial Services

O ROLI MACIERZY KOWARIANCJI W LINIOWEJ FUNKCJI DYSKRYMINACYJNEJ

Streszczenie: Konstrukcja znanej i często wykorzystywanej w praktyce liniowej funkcji dyskryminacyjnej oparta jest na macierzy kowariancji wewnątrzklasowej. W klasycznym przypadku analizy dyskryminacyjnej, gdy występują dwie populacje, liniowa funkcja dyskryminacyjna jest w pewnym sensie optymalna, gdy obie populacje mają tę samą macierz kowariancji. W artykule podniesiony został problem wykorzystania macierzy kowariancji całkowitej w funkcji dyskryminacyjnej. W wyniku przeprowadzonych analiz okazuje się bowiem, że jakość klasyfikacji przeprowadzonych przy użyciu macierzy kowariancji zarówno całkowitej, jak i wewnątrzklasowej jest identyczna. To ciekawe zjawisko stało się inspiracją do znalezienia przyczyn oraz interpretacji geometrycznej owej reguły.

Słowa kluczowe: analiza dyskryminacyjna, credit scoring, wektory własne, macierz kowariancji.

Problematyka związana z zagadnieniem klasyfikacji obserwacji do jednej ze znanych populacji ma swoje bogate odzwierciedlenie w literaturze z obszaru wielowymiarowej analizy danych. W toku licznych badań przeprowadzonych nad efektywnością różnorodnych algorytmów umożliwiających znalezienie reguły klasyfikacyjnej, która z jednej strony minimalizowałaby błąd klasyfikacji w ramach próby uczącej, a z drugiej strony zachowywała wysoką predyktywność w zastosowaniach praktycznych, okazało się, że podstawowe metody często pozwalają uzyskać wyniki nie gorsze od metod dużo bardziej zaawansowanych. Wnioski wyciągnięte z badań pokrywają się również z doświadczeniami tych osób, które budowały modele statystyczne, a następnie stosowały metody dyskryminacyjne w finansach w obszarze badania wiarygodności kredytowej klientów indywidualnych. Stosunkowo najczęściej wykorzystywana przez bankowców metoda oceny wiarygodności kredytowej oparta jest na liniowej funkcji dyskryminacyjnej Fishera. Prostota metody odznacza się duża łatwością w interpretacji parametrów modelu, co z pewnością nie pozostaje bez znaczenia dla praktyków.

Stosunkowo dobrze znana i szeroko opisana w literaturze liniowa funkcja dyskryminacyjna Fishera wykorzystuje w swojej budowie wewnątrzklasową macierz kowariancji cech. Takie postępowanie przy budowie modelu zapewnia nam jego optymalność rozumianą w sensie przedstawionym w dalszej części artykułu. W drodze licznie przeprowadzanych eksperymentów nad modelami dyskryminacyjnymi pojawiła się pewna hipoteza co do zasadności wykorzystywania w procesie budowy modelu liniowej funkcji dyskryminacyjnej Fishera macierzy kowariancji wewnątrzklasowej. Stąd postanowiono zbadać wpływ na jakość modelu podstawienia do wzoru Fishera w miejsce macierzy kowariancji wewnątrzklasowej macierzy kowariancji całkowitej wyznaczonej dla obu populacji łącznie. Okazuje się bowiem, że parametry modelu uzyskane w ten sposób różnią się od tych uzyskanych tradycyjnie. Co ciekawe, w obu przypadkach jakość klasyfikacji zawsze pozostaje identyczna. Zjawisko to stało się przyczyną poszukiwania jego wytłumaczenia oraz powodem odnalezienia jego geometrycznej interpretacji.

R.A. Fisher w pracy [1936] zaproponował metodę analizy dyskryminacyjnej, w której wektor parametrów wyznaczany był na podstawie poniższego wzoru:

$$\hat{\lambda} = S^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \tag{1}$$

gdzie $\bar{\boldsymbol{x}}_1$ oraz $\bar{\boldsymbol{x}}_2$ to wektory wartości średnich wyznaczane dla obu populacji, natomiast macierz S nie jest „zwykłą” macierzą kowariancji liczoną jako:

$$S^{(C)} = \frac{1}{n}\left[\sum_i (x_i - \bar{\boldsymbol{x}})(x_i - \bar{\boldsymbol{x}})^T\right], \tag{2}$$

lecz macierzą wariancji wewnątrzklasowej liczoną według wzoru:

$$S^{(W)} = \frac{1}{n}\left[\sum_{j=1}^{K}\sum_{c_i \in C_j} (x_i - \bar{\boldsymbol{x}}_j)(x_i - \bar{\boldsymbol{x}}_j)^T\right], \tag{3}$$

czy też równoważnie, jak podaje Maddala w pracy [Maddala 1994]:

$$S^{(W)} = \frac{1}{n_1 + n_2}\left[\sum_i (x_{1i} - \bar{\boldsymbol{x}}_1)(x_{1i} - \bar{\boldsymbol{x}}_1)^T + \sum_i (x_{2i} - \bar{\boldsymbol{x}}_2)(x_{2i} - \bar{\boldsymbol{x}}_2)^T\right]. \tag{4}$$

Warto zauważyć, że z faktu występowania takich samych macierzy kowariancji w obu populacjach (bo rozważamy przypadek dwóch populacji), tj. $\Sigma_1 = \Sigma_2$, nie wynika, że macierz kowariancji liczona dla wszystkich obserwacji Σ jest równa macierzom Σ_1 oraz Σ_2. Jest tak jedynie w trywialnym przypadku, gdy dwie populacje pochodzą z tego samego rozkładu, co łatwo udowodnić dla przypadku populacji o rozkładzie jednowymiarowym.

Tak więc w zależności od tego, czy do wzoru (1) podstawimy macierz liczoną według wzoru (2) czy też według wzoru (3), otrzymamy za każdym razem inny wektor parametrów $\hat{\lambda}$. Można sobie zatem zadać pytanie, który z uzyskanych wektorów parametrów jest lepszy, tj. który pozwala lepiej klasyfikować obserwacje. Aby odpowiedzieć na to pytanie, należy wyjść od znanego związku między macierzami kowariancji wewnąrzgrupowej, całkowitej i międzygrupowej postaci:

$$S^{(C)} = S^{(W)} + S^{(M)}.$$

W zaproponowanej przez Fishera metodzie parametry modelu analizy dyskryminacyjnej wyznacza się, maksymalizując wyrażenie:

$$K^{(1)} = \frac{\lambda^T (\overline{x}_1 - \overline{x}_2)^2}{\lambda^T S^{(W)} \lambda} \to \max, \tag{5}$$

gdzie w liczniku mamy wariancję międzygrupową, mianownik zaś to wariancja wewnątrzgrupowa. Licznik można zatem zapisać (podobnie jak mianownik) jako $\lambda^T S^{(M)} \lambda$, gdzie $S^{(M)}$ to macierz kowariancji międzyklasowej:

$$K^{(1)} = \frac{\lambda^T S^{(M)} \lambda}{\lambda^T S^{(W)} \lambda} \to \max \tag{6}$$

lub, co jest równoważne, minimalizując odwrotność:

$$\frac{1}{K^{(1)}} = \frac{\lambda^T S^{(W)} \lambda}{\lambda^T S^{(M)} \lambda} \to \min. \tag{7}$$

Zmieniając kryterium (6) w taki sposób, że macierz kowariancji wewnątrzgrupowej $S^{(W)}$ zastąpimy macierzą kowariancji całkowitej $S^{(C)}$, otrzymamy:

$$K^{(2)} = \frac{\lambda^T S^{(M)} \lambda}{\lambda^T S^{(C)} \lambda} = \frac{\lambda^T S^{(M)} \lambda}{\lambda^T (S^{(M)} + S^{(W)}) \lambda} = \frac{\lambda^T S^{(M)} \lambda}{\lambda^T S^{(M)} \lambda + \lambda^T S^{(W)} \lambda} \to \max. \tag{8}$$

Szukając wektora rozwiązań λ, można powyższe kryterium (8) zapisać w postaci minimalizacji odwrotności, tj.

$$\frac{1}{K^{(2)}} = \frac{\lambda^T S^{(M)} \lambda + \lambda^T S^{(W)} \lambda}{\lambda^T S^{(M)} \lambda} \to \min, \tag{9}$$

można tu zauważyć, że

$$\frac{1}{K^{(2)}} = \frac{\lambda^T S^{(M)} \lambda + \lambda^T S^{(W)} \lambda}{\lambda^T S^{(M)} \lambda} = \frac{\lambda^T S^{(M)} \lambda + \lambda^T S^{(W)} \lambda}{\lambda^T S^{(M)} \lambda} = 1 + \frac{\lambda^T S^{(W)} \lambda}{\lambda^T S^{(M)} \lambda} = 1 + \frac{1}{K^{(1)}}. \tag{10}$$

Z powyższej równości wynika, że minimalizowana funkcja (7) różni się od funkcji (9) jedynie o stałą, co znacznie upraszcza dalsze wnioskowanie co do znalezienia wektora rozwiązań w obu rozważanych wariantach. Zatem okazuje się, że nie jest istotne, którą macierz weźmiemy do analizy $S^{(W)}$ czy też $S^{(C)}$, bo w rezultacie uzyskany wektor $\hat{\lambda}$ będzie spełniał warunek minimalizacji funkcji zarówno (7), jak i (9).

Analizując wektory $\hat{\lambda}$ uzyskane według dwóch przedstawionych sposobów, przekonujemy się, się że ich wartości składowe są różne, choć różnica obu uzyskanych wektorów polega jedynie na różnicy w ich długościach, podczas gdy ich kierunki pozostają identyczne. Jak to zostało przedstawione powyżej, wektory rozwiązań uzyskane dla różnych macierzy kowariancji korespondują z ekstremum obu przedstawionych kryteriów, tj. K1 (6) oraz K2 (8). Stąd zatem wniosek, iż ekstremum obu funkcji K1 i K2 jest niewrażliwe na długość wektora rozwiązań λ.

Co ciekawe, korzystając z ogólniejszego zapisu kryterium:

$$K^{(1)} = \frac{\lambda^T S^{(M)} \lambda}{\lambda^T S^{(W)} \lambda} \rightarrow \max,$$

można wykazać, że λ jest wektorem własnym niesymetrycznej macierzy A postaci:

$$A = S^{(W)-1} S^{(M)}.$$

Zależność ta doskonale tłumaczy, dlaczego uzyskane dwa „różne” wektory mogą być rozwiązaniami równoważnymi. Otóż, gdy rozwiązaniem jest wektor własny macierzy A, to mnożąc go przez skalar, dokonujemy jedynie zmiany jego długości, podczas gdy dalej pozostaje on wektorem własnym macierzy A, w związku z czym dalej jest on poprawnym rozwiązaniem. Taki właśnie przypadek obserwujemy przez wykorzystanie we wzorze Fishera macierzy kowariancji całkowitej, gdy uzyskany wektor rozwiązań różni się jedynie swoją długością od wektora rozwiązań uzyskanego droga tradycyjną.

Celem przedstawionych powyżej rozważań o charakterze teoretycznym było wyjaśnienie interesującego zjawiska polegającego na otrzymywaniu analogicznych wyników analizy dyskryminacyjnej w kontekście wykorzystania obu macierzy kowariancyjnych. Wyjaśnienie przyczyn relacji pomiędzy uzyskanymi wektorami parametrów pozwala również uprościć proces budowy modelu dyskryminacyjnego. Łatwiej bowiem jest obliczyć jedną macierz kowariancji całkowitej, której implementacja znajduje się w każdym oprogramowaniu statystycznym, jak również w popularnym Excelu, a następnie podstawić ją do wzoru Fishera (co uczyniłem), aniżeli wyznaczać dwie odrębne macierze kowariancji wewnątrzklasowej dla każdej z populacji z osobna i wyznaczać trzecią macierz, którą według procedury wykorzystujemy we wzorze Fishera. W ten sposób możliwe staje się skrócenie czasu przygotowania obliczeń pośrednich, co jest szczególnie ważne w przypadku modeli o dużej liczbie obserwacji stosowanych w praktyce. Ponadto cały model nabiera prostoty oraz elegancji, co nie pozostaje bez znaczenia w kontekście odkrywania piękna matematyki.

Literatura

Fisher R.A. (1936), *The use of multiple measurements in taxonomic problems*, „Annals of Eugenics”, nr 7.

Jajuga K. (1990), *Statystyczna teoria rozpoznawania obrazów,* PWN, Warszawa.

Maddala G.S. (1994), *Limited-dependent and qualitative variables in econometrics* Cambridge Univ. Press.

Seber G.A.F. (1984), *Multivariate Observations*, Wiley, New York.

THE ROLE OF THE COVARIANCE MATRIX IN A LINEAR DISCRIMINANT FUNCTION

Summary: The construction of the well-known and often used in practice linear discriminant function is based on the interclass covariance matrix. In a classic case of discriminant analysis, where there are two populations, linear discriminant function is optimal when both populations have the same covariance matrix.In this article the author describes a problem of using the total covariance matrix calculated for all observations in the discriminant function. As a result of the analysis it is clear that the quality of the classification carried out by using the total covariance matrix and interclass is identical. This interesting phenomenon has become an inspiration to find the causes and the geometric interpretation of this rule.

Key words: discriminant analysis, credit scoring, eigenvectors, covariance matrix.