Лекція 9. Дискримінантна функція Фішера

Припустимо, що ми маємо дві генеральні сукупності. Нехай $X_1, X_2, \ldots, X_{n_1}$ — це n_1 спостережень з першої генеральної сукупності, а $X_{n_1+1}, X_{n_1+2}, \ldots, X_{n_1+n_2}$ — n_2 спостережень з другої генеральної сукупності. Вважатимемо, що $X_1, X_2, \ldots, X_{n_1}$, $X_{n_1+1}, X_{n_1+2}, \ldots, X_{n_1+n_2}$ — вектори чисел в просторі розмірності p. Дискримінантний метод Фішера полягає у проектуванні цих векторів із простору розмірності p на числову пряму за допомогою лінійної функції $l(X) = w^T X$ й *відокремленні двох генеральних сукупностей якомога далі одна від одної за допомогою вектора* w *з простору розмірності* p.

1. Дискримінантний метод Фішера.

Необхідно знайти вектор $\hat{w}$, що максимізує функцію критерію $J(w)$, де

$$J(w) = \frac{\overline{Y}_1 - \overline{Y}_2}{S_Y},$$

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n_1} Y_i}{n_1}, \quad \overline{Y}_2 = \frac{\sum_{i=n_1+1}^{n_1+n_2} Y_i}{n_2}, \quad S_Y^2 = \frac{\sum_{i=1}^{n_1}(Y_i - \overline{Y}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2}(Y_i - \overline{Y}_2)^2}{n_1 + n_2 - 2},$$

$$Y_i = w^T X_i, \quad i = 1, 2, \ldots, n_1 + n_2$$

З інтуїтивної точки зору, функція критерію $J(w) = \frac{\overline{Y}_1 - \overline{Y}_2}{S_Y}$ оцінює різницю між середніми проекцій $\overline{Y}_1 - \overline{Y}_2$ відносно до стандартного відхилення S_Y. Якщо проекції $Y_1, Y_2, \ldots, Y_{n_1}$ і $Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_{n_1+n_2}$ можна відокремити повністю, то величина $\left|\overline{Y}_1 - \overline{Y}_2\right|$ повинна бути великою відносно до середнього відхилення S_Y.

Теорема. Вектор $\hat{w}$, що максимізує величину $J(w) = \left|\frac{\overline{Y}_1 - \overline{Y}_2}{S_Y}\right|$, має вигляд $S_W^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$,

де

$$S_W = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}, \quad S_1 = \frac{\sum_{i=1}^{n_1}\left(X_i - \overline{X}_1\right)\left(X_i - \overline{X}_1\right)^T}{n_1 - 1}, \quad S_2 = \frac{\sum_{i=n_1+1}^{n_1+n_2}\left(X_i - \overline{X}_2\right)\left(X_i - \overline{X}_2\right)^T}{n_2 - 1},$$

$$\overline{X}_1 = \frac{\sum_{i=1}^{n_1} X_i}{n_1} \quad \text{і} \quad \overline{X}_2 = \frac{\sum_{i=n_1+1}^{n_1+n_2} X_i}{n_2}.$$

Доведення.

$$\overline{Y}_1 = \frac{\sum_{i=1}^{n_1} Y_i}{n_1} = \frac{\sum_{i=1}^{n_1} w^T X_i}{n_1} = w^T \left(\frac{\sum_{i=1}^{n_1} X_i}{n_1}\right) = w^T \overline{X}_1.$$

Аналогічно, $\overline{Y}_2 = w^T \overline{X}_2$. Отже,

$$\sum_{i=1}^{n_1}(Y_i - \overline{Y}_1)^2 = \sum_{i=1}^{n_1}\left(w^T X_i - w^T \overline{X}_1\right)^2 = \sum_{i=1}^{n_1}\left(w^T X_i - w^T \overline{X}_1\right)\left(w^T X_i - w^T \overline{X}_1\right)^T$$

$$= \sum_{i=1}^{n_1} w^T \left(X_i - \overline{X}_1\right)\left(X_i - \overline{X}_1\right)^T w = w^T \left[\sum_{i=1}^{n_1}\left(X_i - \overline{X}_1\right)\left(X_i - \overline{X}_1\right)^T\right] w.$$

Крім того,

$$\sum_{i=n_1+1}^{n_1+n_2}\left(Y_i-\bar{Y}_2\right)^2 = w^T\left[\sum_{i=n_1+1}^{n_1+n_2}\left(X_i-\bar{X}_2\right)\left(X_i-\bar{X}_2\right)^T\right]w$$

Таким чином,

$$S_Y^2=\frac{\sum_{i=1}^{n_1}\left(Y_i-\bar{Y}_1\right)^2+\sum_{i=n_1+1}^{n_1+n_2}\left(Y_i-\bar{Y}_2\right)^2}{n_1+n_2-2}=\frac{w^T\left[\sum_{i=1}^{n_1}\left(X_i-\bar{X}_1\right)\left(X_i-\bar{X}_1\right)^T\right]w+w^T\left[\sum_{i=n_1+1}^{n_1+n_2}\left(X_i-\bar{X}_2\right)\left(X_i-\bar{X}_2\right)^T\right]w}{n_1+n_2-2}=$$

$$=w^T\left[\frac{\sum_{i=1}^{n_1}\left(X_i-\bar{X}_1\right)\left(X_i-\bar{X}_1\right)^T+\sum_{i=n_1+1}^{n_1+n_2}\left(X_i-\bar{X}_2\right)\left(X_i-\bar{X}_2\right)^T}{n_1+n_2-2}\right]w==w^T\left[\frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}\right]w=w^TS_Ww\,.$$

Отже,

$$J(w)=\frac{\bar{Y}_1-\bar{Y}_2}{S_Y}=\frac{w^T\left(\bar{X}_1-\bar{X}_2\right)}{\sqrt{w^TS_Ww}}$$

Вектор $\hat{w}$ можна знайти, розв'язавши рівняння

$$\frac{\partial J(w)}{\partial w}=\frac{\left(\bar{X}_1-\bar{X}_2\right)}{\sqrt{w^TS_Ww}}-\frac{1}{2}w^T\left(\bar{X}_1-\bar{X}_2\right)\frac{2S_Ww}{\left(w^TS_Ww\right)^{3/2}}=0\,.$$

Подальші перетворення приводять до рівняння

$$\bar{X}_1-\bar{X}_2=\left[\frac{w^T\left(\bar{X}_1-\bar{X}_2\right)}{w^TS_Ww}\right]S_Ww\,.$$

Множачи обидві частині цього рівняння на матрицю S_W^{-1}

$$S_W^{-1}\left(\bar{X}_1-\bar{X}_2\right)=\left[\frac{w^T\left(\bar{X}_1-\bar{X}_2\right)}{w^TS_Ww}\right]w\,,$$

Оскільки $\frac{w^T(\bar{X}_1-\bar{X}_2)}{w^TS_Ww}$ — дійсне число, маємо

$$\hat{w}=cS_W^{-1}\left(\bar{X}_1-\bar{X}_2\right),$$

де c — деяка константа.

Матриця S_W називається *матрицею розкиду всередині класу*.

Матриця $S_B=\left(\bar{X}_1-\bar{X}_2\right)\left(\bar{X}_1-\bar{X}_2\right)^T$ називається *матрицею розкиду між класами*.

За допомогою матриць S_W і S_B функцію критерію в методі Фішера можна сформулювати за допомогою *узагальненого відношення Релея*:

$$J\left(w\right)=\frac{w^TS_Bw}{w^TS_Ww}\,.$$

2. Классифікація.

Припустимо, що ми маємо спостереження X_0. Спираючись на дискримінантну функцію $l(X)=\hat{w}^T X$, яку ми знайшли вище, ми можемо віднести це спостереження до певного класу.

Вирішальне правило. *Спостереження* X_0 *належить до генеральної сукупності 1, якщо*

$$\hat{Y}_0=\hat{w}^T X_0=\left(\overline{X}_1-\overline{X}_2\right)^T S_W^{-1} X_0 \geq \frac{1}{2}\hat{w}^T(\overline{X}_1+\overline{X}_2)=\frac{1}{2}(\overline{Y}_1+\overline{Y}_2)=\frac{1}{2}\left(\overline{X}_1-\overline{X}_2\right)^t S_W^{-1}\left(\overline{X}_1+\overline{X}_2\right).$$

У супротивному випадку

$$\hat{Y}_0=(\overline{X}_1-\overline{X}_2)^T S_W^{-1} X_0<\frac{1}{2}\left(\overline{X}_1-\overline{X}_2\right)^T S_W^{-1}\left(\overline{X}_1+\overline{X}_2\right)$$

і спостереження X_0 *належить до генеральної сукупності 2.*

Якщо $\hat{Y}_0$ *менше* $\frac{\overline{Y}_1+\overline{Y}_2}{2}$ *(ближче до* $\overline{Y}_1$*), то відносимо* X_0 *до генеральної сукупності 1 і навпаки.*

Зауваження. Значне відділення не означає доброї класификації. З іншого боку, якщо відділення не є значним, то здійснювати класифікацію немає сенсу.

Приклад 9.1.

$$\overline{x}_1=\begin{bmatrix}-0.0065\\-0.0390\end{bmatrix},\ \overline{x}_2=\begin{bmatrix}-0.2483\\0.0262\end{bmatrix},\ S_W^{-1}=\begin{bmatrix}131.158 & -90.423\\-90.423 & 108.147\end{bmatrix}.$$

Then,

$$\hat{w}=S_W^{-1}\left(\overline{x}_1-\overline{x}_2\right)=\begin{bmatrix}131.158 & -90.423\\-90.423 & 108.147\end{bmatrix}\left(\begin{bmatrix}-0.0065\\-0.0390\end{bmatrix}-\begin{bmatrix}-0.2483\\0.0262\end{bmatrix}\right)=\begin{bmatrix}37.61\\-28.92\end{bmatrix}$$

$$\hat{y}=\hat{w}^T x=37.61x_1-28.92x_2\ .$$

Отже,

$$\overline{y}_1=\hat{w}^T\overline{x}_1=\begin{bmatrix}37.61 & -28.92\end{bmatrix}\begin{bmatrix}-0.0065\\-0.0390\end{bmatrix}=0.88,$$

$$\overline{y}_2=\hat{w}^T\overline{x}_2=\begin{bmatrix}37.61 & -28.92\end{bmatrix}\begin{bmatrix}-0.2483\\0.0262\end{bmatrix}=-10.10\ .$$

$$\frac{\overline{y}_1+\overline{y}_2}{2}=\frac{0.88-10.10}{2}=-4.61$$

Припустимо, що маємо нове спостереження

$$x_0=\begin{bmatrix}-0.210\\-0.044\end{bmatrix}.$$

Тоді

$$\hat{y}_0=\hat{w}^T x_0=\begin{bmatrix}37.61 & -28.92\end{bmatrix}\begin{bmatrix}-0.210\\-0.044\end{bmatrix}=-6.62<-4.61=\frac{\overline{y}_1+\overline{y}_2}{2}\ .$$

Отже, спостереження належить до другої генеральної сукупності.

1. Дуда Р., Харт П. Разпознавание образов и анализ сцен. — М.: Мир, 1976.