МЕТОД ЗА ПОСТРОЯВАНЕ НА КВАДРАТИЧНА РАЗДЕЛЯЩА ФУНКЦИЯ В УСЛОВИЯТА НА ОГРАНИЧЕНА СТАТИСТИКА

Тянев Д. С.

Увод

Необходимостта от многомерна квадратична функция възниква при решаване на задачи в теорията на разпознаване на образи. Най-често тази функция се използва като модел, представящ отделните състояния на изследвания обект. Квадратичната функция се търси в пространството на признаковия модел на обекта, въз основа на статистическата извадка, събрана за него. Статистическите извадки на признаковия модел могат да бъдат няколко и да се отнасят за отделните състояния на обекта. Според вероятностния смисъл на наблюденията (наричани още образи) събрани за обекта, квадратичната функция е оптимална [1]. Тази функция е подходяща за случая, когато е познато само едно от състоянията на обекта, благодарение на това, че тя е затворена функция.

Известни са два основни подхода при построяване на квадратичната функция – статистически и детерминиран. Методите на първия подход изискват познаването на законите на разпределение на образите, а методите на втория се основават на най-вероятната точка в пространството на признаците, представяща даденото състояние. В крайна сметка, при определени условия, резултатите са аналогични [2].

За съжаление редица реални обекти не могат да осигурят условията за приложение на тези методи. Обикновено това са сложни обекти, чийто признаков модел е с голяма размерност, намиращи се в непрекъсната експлоатация, непригодни за активни експерименти, рядко случващи се във времето състояния и притежаващи строга дисциплина на обслужване. В такива условия обемът на статистическото описание на отделните състояния на обекта е крайно ограничено – обикновено няколко десетки наблюдения, които са събрани за значителен временен интервал на експлоатация. В тези условия (наричани тук *ограничена статистика*) вероятностните оценки са ненадеждни.

Постановка на подхода

Предложеният тук подход се основава на разбирането ни, че избраната функция, разглеждана като затворена повърхнина в пространството на признаците, може да се търси като обвиваща статистическата извадка елипсоидална повърхнина. При това тя трябва да обхваща множеството образи възможно най-плътно. Търсеният хиперелипсоид следва да бъде ориентиран в системата от координатни оси на максималната разпръстнатост на образите. Намирането на векторния базис на максималната разпръстнатост обикновено се извършва чрез оценка на разпръстнатостта спрямо най-вероятната точка, която най-често се оценява като център на тежестта на даденото множество образи. Като се имат предвид условията на ограничена статистика и ненадеждността на тази оценка, както и взаимната зависимост между векторния базис на максималната разпръстнатост и точката за центриране на функцията, тук се предлага:

- Векторният базис да бъде намерен без предварително центриране на извадката, чрез множеството на възможните хорди;
- Точката за центриране на квадратичната функция да се определи като точка на геометрична симетрия в пространството на максималната разпръстнатост.

Метод

Като критерий за максимална разпръснатост определяме сумата от дължините на проекциите на възможните в множеството образи хорди $\boldsymbol{p}_l$:

$$\boldsymbol{p}_l = \boldsymbol{x}_q - \boldsymbol{x}_r \;; \quad q,r = \overline{1,m} \;; \quad q < r \;; \quad l = \overline{1,n} \;; \quad n = \binom{m}{2} , \qquad \textbf{(1)}$$

т.е. максималната разпръстнатост на образите в признаковото пространство е ориентирана по направлението на единичния вектор $\boldsymbol{s}$, ако сумата

$$\sum_{l=1}^{n} \left| (\boldsymbol{p}_l, \boldsymbol{x}) \right| \tag{2}$$

достига своя максимум при $\boldsymbol{x}=\boldsymbol{s}$.

Очевидно тази сума има екстремум там където го има и сумата

$$\sum_{l=1}^{n} \left| (\boldsymbol{p}_l, \boldsymbol{x}) \right|^2 \quad , \tag{3}$$

която изразява нормата на матрицата $\boldsymbol{P.x}$, т.е.

$$\sum_{l=1}^{n} \left| (\boldsymbol{p}_l, \boldsymbol{x}) \right|^2 = \|\boldsymbol{P.x}\| \quad ; \qquad \boldsymbol{P}(n,k) = \begin{bmatrix} \boldsymbol{p}_1^T \\ \boldsymbol{p}_2^T \\ \cdots \\ \boldsymbol{p}_n^T \end{bmatrix} .$$

Чрез метода на Лагранж за намиране на условен екстремум с функция на условието

$$\omega = \sum_{j=1}^{k} x_j^2 \quad ; \qquad \boldsymbol{x} = [\, x_1\, x_2 \ldots x_k \,]^t \tag{4}$$

получаваме, че съответният множител на Лагранж – λ_1 е собствената стойност на нормалната матрица $\boldsymbol{P}^T\!.\boldsymbol{P}$, а $\boldsymbol{s}$ е съответен на λ_1 собствен вектор. При това за да бъде нормата

$$\|\boldsymbol{P.s}\| = \lambda_1.\|\boldsymbol{s}\| = \lambda_1$$

максимална, то множителят λ_1 трябва да бъде най-голямата собствена стойност. Така определеното направление на максималната разпрътнатост не е обвързано с предварително центриране на изходните данни.

След определянето по този начин на първия от търсените базисни вектори – $\boldsymbol{s}_1$, множеството от хорди $\boldsymbol{P}$ следва да се проектира в ортогоналното му допълнение. Ако означим ортогоналното допълнение на $\boldsymbol{s}_1$ с матрицата $\boldsymbol{V}^{(k-1)}$, която е с размери (*k*,*k*-1), то проекцията на $\boldsymbol{P}$ в това подпространство се намира така:

$$\boldsymbol{P}.\boldsymbol{V}^{(k-1)} = \boldsymbol{P1} \quad . \tag{5}$$

В това подпространство за определяне направлението на максималната разпрътнатост се прилага същия критерий. От симетричността на дясната нормална матрица $\boldsymbol{P}^T\!.\boldsymbol{P}$ обаче следва, че така определеният чрез матрицата $\boldsymbol{P1}^T\!.\boldsymbol{P1}$ втори по ред вектор $\boldsymbol{s}_2$, – е също собствен вектор на $\boldsymbol{P}^T\!.\boldsymbol{P}$. Следователно, ако всички собствени стойности на тази матрица са различни, ортогоналният векторен базис, определен от съответно подредените й собствени вектори, е единственият базис, който удовлетворява въведения критерий за максимална разпрътнатост.

Случай на кратни собствени стойности

В случай, че дадена собствена стойност λ на матрицата $\boldsymbol{P}^T\!.\boldsymbol{P}$ се получи с кратност r, $r \geq 2$, то от безкрайно многото ортогонални помежду си системи от r собствени вектори, съответстващи на λ , е необходимо да бъде избрана една конкретна система. Този избор не може да се осъществи въз основа на използвания до момента критерий, тъй като неговата стойност върху всяко направление в това r-мерно подпространство, е една и съща.

Отчитайки естественото ни изискване търсената повърхнина да обхваща възможно по-плътно образите от статистическата извадка, избираме поредната ос (първата за подпространството) да бъде ориентирана в направлението на най-дългата хорда $\boldsymbol{p}_{l_{max}}^{(r)}$ – (или на която и да е от тях, ако са няколко). В случая, когато тези хорди са няколко, изборът на направлението може да се прецизира допълнително чрез усредняването им. Този нов критерий отчита факта, че центърът на симетрия, който трябва да се определи в последствие, е в пряка зависимост от избора на главните направления на разпрътнатостта.

Следващата по ред ос се определя по същият начин след проектиране на множеството хорди в подпространството, ортогонално допълващо вектора $\boldsymbol{p}_{l_{max}}^{(r)}$. Този критерий за избор на направленията на максималната разпрътнатост се прилага до изчерпване на размерността на подпространството или най-много $(r\text{-}1)$ стъпки. Същността на тази последователност е следната: ако първите ρ на брой собствени стойности на матрицата $\boldsymbol{P}^T.\boldsymbol{P}$ са различни, то еднозначно определени са и първите собствени вектори : $(\boldsymbol{s}_1,\ \boldsymbol{s}_2,\ \boldsymbol{s}_3\ ,\ \dots\ ,\ \boldsymbol{s}_\rho)$. Ако следващата собствена стойност λ е с кратност r , то следва да се намери векторен базис $\boldsymbol{V}^{(r)}(k,r)$, ортогонален спрямо базиса $\boldsymbol{V}^{(\rho)}$. Търсеното множество вектори се получава като общо решение на линейната система

$$(\,\boldsymbol{P}^T.\boldsymbol{P}-\lambda.\boldsymbol{E}\,).\boldsymbol{x}=0\ , \qquad \textbf{(6)}$$

от която формираме една фундаментална система решения, получена стандартно във вида:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & 1 \\ a_{r+1} & a_{r+1} & \cdots & a_{r+1} \\ a_k & a_k & \cdots & a_k \end{bmatrix} \qquad \textbf{(7)}$$

фундаменталната система решения (7) – $\boldsymbol{G}(k,r)$, се ортогонализира с помощта на метода на Грам-Шмид, след което се нормира и се получава окончателният вид на векторния базис, който търсим – $\boldsymbol{V}^{(r)}(k,r)$. След това се извършва проектиране на множеството хорди в този базис:

$$\boldsymbol{P}(\,n,k\,).\boldsymbol{V}^{(\,r\,)}(\,k,r\,)=\boldsymbol{P1}(\,n,k\,)$$

и в проекцията $\boldsymbol{P1}$ се намира най-дългата хорда $\boldsymbol{p1}_{l_{max}}^{(r)}$, от която след нормиране на представящия я вектор, получаваме вектора на поредната ос:

$$\boldsymbol{s}_{\rho+1}^{(\,r\,)}=\frac{\boldsymbol{p1}_{l_{max}}^{(\,r\,)}}{\left\|\boldsymbol{p1}_{l_{max}}^{(\,r\,)}\right\|}\ ,$$

който в изходното пространство ще има вида:

$$\boldsymbol{s}_{\rho+1}=\boldsymbol{V}^{(\,r\,)}.\boldsymbol{s}_{\rho+1}^{(\,r\,)}\ .$$

Следващата в подпространството ос – $\boldsymbol{s}_{\rho+2}^{(\,r\,)}$ се търси като ортогонална на вече намерената $\boldsymbol{s}_{\rho+1}^{(\,r\,)}$, удовлетворяваща уравнението на скаларното произведение:

$$\left(\boldsymbol{s}_{\rho+1}^{(\,r\,)},\boldsymbol{y}^{(\,r\,)}\right)=0\ . \qquad \textbf{(8)}$$

Координатите на векторите $\boldsymbol{y}$ се получават като общо решение на уравнението (8) като една фундаментална система решения от вида:

$$\begin{bmatrix} \dfrac{-s^{(r)}_{\rho+1,2}}{s^{(r)}_{\rho+1,1}} & \dfrac{-s^{(r)}_{\rho+1,3}}{s^{(r)}_{\rho+1,1}} & \cdots & \dfrac{-s^{(r)}_{\rho+1,r}}{s^{(r)}_{\rho+1,1}} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & & 1 \end{bmatrix} ,$$

която с помощта на метода на Грам-Шмид се ортогонализира и след нормиране се получава търсеният векторен базис $\boldsymbol{V}^{(r-1)}(r,r-1)$. Следва проектирането

$$\boldsymbol{P1}(n,r).\boldsymbol{V}^{(r-1)} = \boldsymbol{P2}(n,r-1) \ ,$$

където в множеството се намира най-дългата хорда $\boldsymbol{p2}^{(r-1)}_{l_{max}}$. Нормира се представящият я вектор и се получава втората по ред ос в подпространството:

$$\boldsymbol{s}^{(r-1)}_{\rho+2} = \frac{\boldsymbol{p2}^{(r-1)}_{l_{max}}}{\left\| \boldsymbol{p2}^{(r-1)}_{l_{max}} \right\|} \ .$$

Тази стъпка се повтаря за определяне на осите $\boldsymbol{s}^{(r-2)}_{\rho+3}$, $\boldsymbol{s}^{(r-3)}_{\rho+4}$ и т.н. до изчерпване на подпространството.

Квадратична функция

След определяне на пълния векторен базис $V=[\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3 , \ \dots \ , \boldsymbol{s}_k]$ на максималната разпръстнатост, който трябва да се разглежда като нова дясно ориентирана координатна система, следва да се намери проекцията на статистическата извадка в тази нова координатна система:

$$\boldsymbol{X}(m,k).\boldsymbol{V} = \boldsymbol{Y}(m,k) \ . \tag{9}$$

Тъй като сега вече множеството $\boldsymbol{Y}$ е ориентирано по направленията на максималната разпръстнатост, могат да се определят координатите на геометричния център на симетрия:

$$O_j = \frac{1}{2}\left(\max_i y_{ij} + \min_i y_{ij} \right) \ ; \quad j = \overline{1,k} \ ; \quad i = \overline{1,m} \ . \tag{10}$$

Като се има предвид, че най-удобен вид за уравнението на квадратичната функция е каноничният, извършваме транслация на координатната система в точка $\boldsymbol{O}$:

$$\boldsymbol{z}_i = \boldsymbol{y}_i - \boldsymbol{O} \ ; \quad i = \overline{1,m} \ , \tag{11}$$

след което се получава окончателният му вид:

$$\sum_{j=1}^{\delta} \frac{z_j^2}{\alpha^2 . \lambda_j} = 1 \ , \tag{12}$$

където $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\delta > 0 \ ; \ \ \delta \leq k$, са получените ненулеви собствени стойности на матрицата $\boldsymbol{P}^T.\boldsymbol{P}$.

Уравнение (12) представлява семейство хиперелипсоидални повърхнини, от което се търси онази, която обхваща всички образи от извадката. Това става чрез конкретизиране на параметъра α:

$$\alpha^2 = \max_i \left(\sum_{j=1}^{\delta} \frac{z_{ij}^2}{\lambda_j} \right) \; ; \quad i = \overline{1,m} \; . \tag{13}$$

Чрез дължините на полуосите $a_j^2 = \alpha^2 . \lambda_j$ можем да изчисляваме обема на хиперелипсоида, който служи като критерий за сравнение [3]. Окончателният вид на квадратичната разделяща функция е следният:

$$d(z) = \sum_{j=1}^{\delta} \frac{z_j^2}{a_j^2} - 1 \; . \tag{14}$$

В изходната координатна система на признаковото пространство функцията има вида:

$$d(\boldsymbol{x}) = \alpha^{-2} (\boldsymbol{x} - \boldsymbol{O}_x)^T . (\boldsymbol{P}^T . \boldsymbol{P})^{-1} . (\boldsymbol{x} - \boldsymbol{O}_x) - 1 \; . \tag{15}$$

Тъй като разполагаме с елементите на ортогоналната декомпозиция на нормалната матрица $\boldsymbol{P}^T . \boldsymbol{P}$ – двойката матрици $(\boldsymbol{V}, \boldsymbol{D})$, стойността на квадратичната форма в уравнение (15) се търси като скаларно произведение на вектор: $\boldsymbol{b}^T . \boldsymbol{A}^{-1} . \boldsymbol{b} = (\boldsymbol{q}, \boldsymbol{q})$. Като имаме предвид, че

$$\boldsymbol{A}^{-1} = \boldsymbol{V} . \boldsymbol{D}^{-1} \boldsymbol{V}^T = \boldsymbol{V} . \boldsymbol{D}^{-1/2} . \boldsymbol{D}^{-1/2} . \boldsymbol{V}^T \quad ,$$

благодарение на това, че $\boldsymbol{D}$ е диагонална матрица, векторът $\boldsymbol{q}$ може да се определи по следния начин:

$$\boldsymbol{q} = \boldsymbol{D}^{-1/2} . \boldsymbol{V}^T . \boldsymbol{b} \; .$$

Крайният вид на квадратичната функция, освободен от матричното обръщане, е

$$d(\boldsymbol{x}) = \alpha^{-2} . \left\| \boldsymbol{D}^{-1/2} . (\boldsymbol{V}^T . \boldsymbol{x} - \boldsymbol{O}) \right\|^2 \; . \tag{16}$$

Заключение

В заключение ще отбележим, че така реализирания подход е изследван както върху моделни, така и върху реални данни. Методът за построяване на квадратичната функция не изисква познаването на статистическия закон за разпределение на данните, нито зависи от обема на извадката. Изчисленията се осъществяват лесно, а сравнителните резултати при симетрични разпределения са аналогични със статистическите. Основно предимство на метода е по-плътното обхващане на извадката в случаите на несиметрично разпределение. Такъв резултат се получава поради лошата апроксимация на такова разпределение със симетрични закони, какъвто е най-често прилаганият $N_x(\boldsymbol{m}, \boldsymbol{C})$.

Литература:

[1]. Фукунага К., *Введение в статистическую теорию распознавания образов,* Москва, Издательство "Наука", 1979.

[2]. Дуда Р., Харт П., *Распознавание образов и анализ сцен,* Москва, Издательство "Наука", 1976.

[3]. Воеводин В., Кузнецов Ю., *Матрицы и вычисления,* Москва, Издательство "Наука", 1984.