ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ДЛЯ РЕШЕНИЯ СПЕЦИАЛЬНОГО КЛАССА ЗАДАЧ РАСПОЗНАВАНИЯ

К.В.Воронцов
Москва

Рассмотрена предварительная обработка данных для задач распознавания, в которых значениями одного из признаков являются разложимые функции двух переменных, заданные на прямоугольной сетке. Построены алгоритмы, позволяющие находить структурные описания этих функций, удовлетворяющие требованиям единственности и устойчивости к погрешностям исходных данных.

Современные методы решения задач распознавания и классификации [1-2] позволяют использовать произвольные множества допустимых значений признаков, в том числе задаваемые путем обобщенного индуктивного определения. Чтобы задать множество таким способом, требуется указать исходные (непроизводные) элементы этого множества и набор операций, с помощью которых можно порождать новые элементы. В этом случае допустимые значения признака удобно представлять в виде структурных описаний, состоящих из описаний отдельных непроизводных элементов и порядка выполнения операций над ними. Как правило такой подход, основанный на явном представлении информации о структуре, позволяет эффективно вычислять оценки близости и экономить память при хранении значений признака. На практике, однако, эти значения могут задаваться в ином виде, и для получения структурных описаний необходимо решать задачу предварительной обработки.

В статье рассматривается частный случай, когда множеством непроизводных значений признака является специальное семейство функций двух переменных, а набор операций состоит из одной операции сложения. Допустимые значения признака, также функции двух переменных, задаются с погрешностями в узлах прямоугольной сетки. Задача предварительной обработки состоит в получении структурного описания заданной сеточной функции.

Естественное требование существования и единственности искомого описания накладывает определенные ограничения на множество непроизводных элементов. В частности, обеспечить единственность можно путем сужения исходного семейства функций за счет привлечения дополнительной информации. Столь же естественно требование устойчивости описания по отношению к погрешностям исходных данных. В данном случае малые погрешности не должны приводить к изменению числа непроизводных элементов в описании.

§1. Построение начального множества описаний

Будем называть непроизводной поверхностью всякую функцию двух переменных вида $f(x)g(y)$, где $f(x)$ и $g(y)$ — произвольные функции, определенные на множествах $\Omega_x \subseteq \mathbb{R}$ и $\Omega_y \subseteq \mathbb{R}$ соответственно, а допустимой поверхностью — всякую функцию двух переменных, представимую в виде суммы конечного числа непроизводных поверхностей.

Рассмотрим следующую задачу. Пусть допустимая поверхность

$$z(x,y) = \sum_{s=1}^{k} f_s(x) g_s(y) \tag{1.1}$$

задана с погрешностями в узлах прямоугольной сетки (ξ_i, η_j), $\xi_i \in \Omega_x$, $\eta_j \in \Omega_y$, $i = 1,\ldots,n$, $j = 1,\ldots,m$ и пусть для определенности $n \le m$. Число непроизводных поверхностей k неизвестно. Требуется найти структурное описание допустимой поверхности, в данном случае определить k и восстановить значения функций $f_s(x)$, $g_s(y)$ в узлах сетки.

Обозначим через $M_{n,m}$ пространство действительных матриц размера $n \times m$ и введем матрицы

$$X = \left[f_s(\xi_i) \right] \in M_{n,k},\ Y = \left[g_s(\eta_j) \right] \in M_{m,k},\ Z = \left[z(\xi_i, \eta_j) \right] \in M_{n,m}.$$

Записав соотношение (1.1) в узлах сетки, получим матричное уравнение

$$XY^T = Z \tag{1.2}$$

относительно неизвестных матриц X и Y с неизвестным числом столбцов k. Для определения k потребуем, чтобы искомое разложение вида (1.1) состояло из минимального числа слагаемых.

Определение. Матрица $Z \in M_{n,m}$ называется *k-разложимой*, если k — наименьшее число, при котором существуют матрицы $X \in M_{n,k}$ и $Y \in M_{m,k}$, удовлетворяющие уравнению $XY^T = Z$. Упорядоченная пара матриц (X,Y) называется *k-разложением Z*.

Множество всех k-разложений Z будем обозначать через M_Z^k.

Лемма 1. *Матрица $Z \in M_{n,m}$, является k-разложимой тогда и только тогда, когда ее ранг равен k. Всякое k-разложение состоит из матриц полного ранга.*

Доказательство. Достаточность. Пусть матрица Z имеет ранг k. Тогда она не может быть q-разложимой при $q < k$. С другой стороны, Z представима в виде произведения $n \times k$- и $k \times m$-матриц (скелетное разложение [3]), причем в качестве столбцов первой из них можно взять k линейно независимых столбцов Z. Матрица Z является k-разложимой.

Необходимость. Пусть (X,Y) — произвольное k-разложение Z. Матрицы X, Y и Z, очевидно, не могут иметь ранг, больший k. Они также не могут иметь ранг, меньший k, так как иначе, по доказанному, матрица Z оказалась бы q-разложимой при $q < k$. Лемма доказана.

Пусть (X,Y) — произвольное k-разложение Z. Тогда матрицы $X^T X$ и $Y^T Y$ невырождены и (1.2) равносильно каждому из равенств

$$Y = Z^T X (X^T X)^{-1}, \quad X = ZY(Y^T Y)^{-1}. \tag{1.3}$$

Лемма 2. *Пусть (X,Y) — какое-либо k-разложение матрицы Z. Тогда множество всех ее k-разложений имеет вид*

$$M_Z^k = \{ (XA, YA^{-1T}) : \ A \in M_{k,k},\ \mathrm{rg} A = k \}.$$

Доказательство. Очевидно, всякая пара матриц указанного вида является k-разложением Z. Верно и обратное: для любого k-

разложения (U,V) найдется невырожденная $k \times k$-матрица A такая, что $U = XA$, $V = YA^{-1T}$. Действительно, используя (1.3), легко проверить истинность этих равенств, если положить $A = Y^T V(V^T V)^{-1}$. Причем A невырождена в силу оценок ранга сверху и снизу: $k = \mathrm{rg}U = \mathrm{rg}XA \le \mathrm{rg}A \le k$. Таким образом, множество указанного вида и множество всех k-разложений матрицы Z совпадают. Лемма доказана.

Непосредственно из лемм 1 и 2 следует

Т е о р е м а 1 . *Если ранг матрицы Z равен k, то она k-разложима и*

$$M_Z^k = \{\ (XA, YA^{-1T}) :\ A \in M_{k,k},\ \mathrm{rg}A = k\},$$

где матрица X составлена из k линейно независимых столбцов Z, матрица Y определяется по формуле $Y = Z^T X(X^T X)^{-1}$.

Таким образом, при данном определении непроизводной поверхности для всякой допустимой поверхности можно указать бесконечно много различных структурных описаний. Для обеспечения единственности описания необходимо вводить дополнительные ограничения на множество непроизводных поверхностей. Эта задача будет рассмотрена в §3.

§2. Вычисление *k*-разложений

Согласно лемме 2, множество всех k-разложений однозначно определяется каким-либо одним его элементом. Поэтому практический интерес представляет задача вычисления произвольного k-разложения данной матрицы. Заметим, что использование с этой целью теоремы 1 нецелесообразно, так как если элементы матрицы Z заданы с погрешностями, то ее ранг может не совпадать с числом непроизводных поверхностей k. Возможный выход состоит в приближенном решении матричного уравнения (1.2) и определении k из условия достаточно точного (в пределах погрешностей Z) выполнения равенства (1.2).

Введем в $M_{n,k}$ евклидову норму $\|A\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2\right)^{1/2}$, где $A = \left[a_{ij}\right]_{n\times m}$.

О п р е д е л е н и е . Матрица Z называется k,δ*-разложимой*, если k— наименьшее число, при котором существует k-разложимая матрица W такая, что $\|Z - W\| \le \delta$. Пара матриц (X,Y) называется k,δ*-разложением Z*, если она является k-разложением W.

Рассмотрим симметрическую неотрицательно определенную матрицу ZZ^T. Все ее собственные числа сосредоточены на отрезке действительной оси $[0,\|Z\|^2]$, а соответствующие им собственные векторы могут быть выбраны так, чтобы они составляли полную ортонормированную систему векторов в R^n. Докажем следующую теорему, примыкающую к известным результатам [4] о сингулярном разложении матриц.

Т е о р е м а 2 . *Пусть заданы матрица* $Z \in M_{n,m}$ *и действительное число* $\delta \ge 0$. *Пусть* $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$ *— все собственные числа матрицы* ZZ^T. *Если k— наименьшее число, при котором выполняется условие*

(2.1) $$\lambda_1 + \ldots + \lambda_k \ge \|Z\|^2 - \delta^2,$$

то матрица Z *является* k,δ *-разложимой, причем множество наилучших в смысле евклидовой нормы* k,δ *-разложений* Z *имеет вид*

$$\{\ (XA, Z^T XA^{-1T}):\ A \in M_{k,k},\ \mathrm{rg}A = k\},$$

где $n \times k$ *-матрица* X *составлена из ортонормированных собственных векторов* ZZ^T*, отвечающих* k *наибольшим собственным числам.*

Доказательство.

Определим функционал $J_q:\ M_{n,q} \times M_{m,q} \to \mathbb{R}$ по формуле

$$J_q(U,V) = \left\|Z - UV^T\right\|^2,$$

полагая при этом $\mathrm{rg}U = \mathrm{rg}V = q$, и найдем пару матриц (U,V), доставляющую J_q глобальный минимум. Дифференцируя J_q по U и V, получим систему нормальных уравнений для данной задачи наименьших квадратов (необходимые условия минимума):

$$\text{(2.2)}\qquad \begin{cases} U^T(Z - UV^T) = 0, \\ (Z - UV^T)V = 0\ . \end{cases}$$

Покажем, что матрица $R = V^T V U^T U$ диагонализируема. Так как матрица $U^T U$ симметрическая и положительно определенная, найдется невырожденная матрица $S \in M_{q,q}$ такая, что $S^{-1T} U^T U S^{-1} = E$. Далее, так как матрица $SV^T VS^T$ симметрическая и невырожденная, найдется ортогональная матрица $T \in M_{q,q}$ такая, что $T(SV^T VS^T)T^{-1} = \Lambda$, где $\Lambda = \mathrm{diag}(d_1,...,d_q)$ — диагональная матрица. Положим $A=TS$. Тогда, как легко проверить, $ARA^{-1} = \Lambda$, то есть R преобразованием подобия с невырожденной преобразующей матрицей A приводится к диагональному виду.

Произведем в системе (2.2) невырожденную замену переменных

$$\text{(2.3)}\qquad U = XA, \quad V = YA^{-1T}.$$

Тогда столбцы новых матриц X и Y будут ортогональны:

$$X^T X = E, \quad Y^T Y = \Lambda,$$

а система (2.2) перейдет в эквивалентную

$$\begin{cases} Y = Z^T X, \\ ZZ^T X = X\Lambda. \end{cases}$$

Второе из полученных уравнений означает, что $d_1,\dots,d_q$ являются собственными числами, а столбцы матрицы X — соответствующими им собственными векторами матрицы ZZ^T.

Для получения не только необходимого, но и достаточного условия минимума вычислим значение функционала J_q^* при U и V, удовлетворяющих (2.2). Так как $J_q(U,V) = J_q(X,Y)$, имеем:

$$J_q^* = \mathrm{tr}(Z - XY^T)^T(Z - XY^T) = \mathrm{tr}ZZ^T - \mathrm{tr}\Lambda = \left\|Z\right\|^2 - (d_1 + ... + d_q).$$

Глобальный минимум функционала $J_q(U,V)$ будет достигаться тогда и только тогда, когда $d_1,\dots,d_q$ будут q наибольшими собственными числами матрицы ZZ^T. Причем, если $q = k$, то в силу условия (2.1)

$$\left\|Z - UV^T\right\|^2 = J_k^* = \left\|Z\right\|^2 - \lambda_1 - \dots - \lambda_k \le \delta^2.$$

Если же $q<k$, то для любых $U \in M_{n,q}$ и $V \in M_{m,q}$ справедливо неравенство

$$\left\| Z - UV^T \right\|^2 \geq J_q^* = \|Z\|^2 - \lambda_1 - \ldots - \lambda_q > \delta^2.$$

Таким образом, матрица Z оказывается k,δ-разложимой. При этом те и только те k,δ-разложения (U,V) минимизируют $J_k(U,V)$, которые преобразованием (2.3) связаны с парой $(X, Z^T X)$, в которой матрица X составлена из ортонормированных собственных векторов ZZ^T , отвечающих k наибольшим собственным числам. Теорема доказана.

Теорема 2 позволяет найти k,δ-разложение Z, если задано δ и известен набор собственных чисел и векторов матрицы ZZ^T. Однако при увеличении размера матрицы Z затраты машинного времени на нахождение спектра существенно возрастают. Поэтому целесообразно рассматривать другие алгоритмы, не требующие явного решения задачи на собственные значения.

Исследуем возможность вычисления k,δ-разложений с помощью итерационного процесса, основанного на соотношениях (1.3).

Введем следующие обозначения. Пусть $X \in M_{n,k}$ — матрица ранга k. Линейную оболочку столбцов матрицы X обозначим через $\mathscr{L}(X)$. Проекционная матрица $P_X = E - X(X^T X)^{-1} X^T$, если рассматривать ее как линейный оператор из $\mathbb{R}^n$ в $\mathbb{R}^n$, ставит в соответствие вектор-столбцу его проекцию на ортогональное дополнение подпространства $\mathscr{L}(X)$. Далее будут использоваться следующие свойства проекционных матриц.

1. Матрица P_X симметрична ($P_X^T = P_X$), идемпотентна ($P_X P_X = P_X$) и вырождена ($P_X X = 0$, ранг P_X равен n-k).

2. Матрица P_X инвариантна относительно выбора базиса в подпространстве $\mathscr{L}(X)$: $P_{XA} = P_X$ для любой невырожденной $A \in M_{k,k}$.

3. Через проекционную матрицу выражается решение следующей задачи наименьших квадратов: $\min\limits_X \|AX - B\|^2 = \|P_A B\|^2$.

Последний факт позволяет определить синус угла $\varphi \in [0, \pi/2]$ между подпространствами $\mathscr{L}(X)$ и $\mathscr{L}(X_0)$, $X_0 \in M_{n,k}$ как максимальное расстояние от единичного вектора одного из подпространств до другого подпространства:

$$\sin\varphi = \max_{a \in M_{k,1}} \frac{\|P_X X_0 a\|}{\|X_0 a\|}. \tag{2.4}$$

Т е о р е м а 3. *Пусть* (X,Y) *— наилучшее в смысле евклидовой нормы* k,δ*-разложение матрицы* $Z \in M_{n,m}$ *ранга n, начальное приближение* $Y_0 \in M_{n,k}$ *удовлетворяет условию* $\operatorname{rg} Y_0^T Y = k$ *,* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ *— все собственные числа матрицы* ZZ^T*. Тогда для итерационного процесса*

$$\begin{cases} X_r = ZY_{r-1}A_r, \\ Y_r = Z^T X_r, \end{cases} \tag{2.5}$$

где матрицы A_r *определяются из условий* $X_r^T X_r = E$*,* $r = 1, 2, \ldots$*, справедлива оценка*

(2.6) $$\left\| Z - X_r X_r^T \right\| \le \delta + \delta\sqrt{k}\mu^{r-1}\ \mathrm{tg}\varphi,$$

где $\mu = \delta^2 \left\| \Lambda^{-1} \right\|$, $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_k)$, φ — *угол между* $\mathscr{L}(X)$ *и* $\mathscr{L}(X_0)$.

Доказательство. Поскольку матрицы Z и Y_0 имеют полный ранг, то и при любом $r > 0$ ранг матриц X_r и Y_r равен k. Следовательно на каждой итерации существует невырожденная верхняя треугольная матрица A_r, определяемая условием $X_r^T X_r = E$. Для ее построения достаточно применить процесс ортогонализации Грама-Шмидта [3,4] к столбцам матрицы ZY_{r-1}.

Путем несложных преобразований легко проверить тождество
$$Z - X_r Y_r^T = P_{X_r} Z = P_{X_r}(Z - XY^T)\left(P_Y - P_Y Y_{r-1}(Y^T Y_{r-1})^{-1} Y^T\right).$$

Введем обозначение $Q_r = P_Y Y_r (Y^T Y_r)^{-1} Y^T$, $r = 0, 1, 2, \ldots$ Используя (2.5), выразим матрицу Q_r через Y_{r-1}:
$$Q_r = (P_Y Z^T Z P_Y) P_Y Y_{r-1} (Y^T Y_{r-1})^{-1} \Lambda^{-1} Y^T.$$

Учитывая, что $Y^T Y = \Lambda$ и $\left\| Z P_Y \right\| \le \delta$, оценим норму Q_r:
$$\left\| Q_r \right\| \le \left\| Z P_Y \right\|^2 \left\| Q_{r-1} \right\| \left\| \Lambda^{-1} \right\| \le \mu \left\| Q_{r-1} \right\|, \ r = 1, 2, \ldots$$

При умножении произвольной матрицы на проекционную ее норма может только уменьшиться, поэтому справедлива оценка
$$\left\| Z - X_r Y_r^T \right\| \le \delta + \delta \left\| Q_{r-1} \right\| \le \delta + \delta\mu^{r-1} \left\| Q_0 \right\|.$$

Оценим теперь норму Q_0. Очевидно, найдутся ортоногональные $m \times k$-матрицы $\tilde{Y}$, $\tilde{Y}_0$ и невырожденные $k \times k$-матрицы A, A_0 такие, что $Y = \tilde{Y} A$ и $Y_0 = \tilde{Y}_0 A_0$. Используя тождество $\left\| Q_0 \right\|^2 = \mathrm{tr}\ Q_0^T Q_0$, получим:
$$\left\| Q_0 \right\|^2 \le \mathrm{tr}\ (\tilde{Y}_0^T \tilde{Y} \tilde{Y}^T \tilde{Y}_0)^{-1} - \mathrm{tr}\ \tilde{Y}^T \tilde{Y} \le k / \lambda - k,$$
где λ — минимальное собственное число матрицы $G = \tilde{Y}_0^T \tilde{Y} \tilde{Y}^T \tilde{Y}_0$. Поскольку матрица G симметрическая и положительно определенная,
$$\lambda = \min_a \frac{a^T G a}{a^T a}.$$

Делая в G подстановку $\tilde{Y}\tilde{Y}^T = E - P_{\tilde{Y}}$, с помощью (2.4) находим:
$$\lambda = 1 - \max_a \frac{\left\| P_Y Y_0 a \right\|^2}{\left\| Y_0 a \right\|^2} = \cos^2 \varphi.$$

Заметим, что $\cos\varphi \ne 0$ в силу невырожденности матрицы $Y_0^T Y$. Таким образом, справедлива оценка
$$\left\| Q_0 \right\| \le \sqrt{\frac{k}{\cos^2 \varphi} - k} = \sqrt{k}\ \mathrm{tg}\varphi,$$
откуда немедленно следует требуемое неравенство. Теорема доказана.

Оценка (2.6) показывает, что при достаточно малой погрешности начальных данных процесс (2.5) позволяет достичь приемлемой точности уже на первых итерациях. В частности, если исходная матрица k-разложима ($\delta = 0$), равенство $Z = X_1 Y_1^T$ выполняется точно для любого Y_0, удовлетворяющего условию $\mathrm{rg}\, Y_0^T Y = k$.

Заметим, что если матрицы A_r верхние треугольные, то для проведения любого числа итераций с q первыми столбцами матриц X и Y не требуется знать остальные k-q столбцов. Этот факт позволяет построить более гибкий алгоритм, в котором итерации по r чередуются с добавлением новых столбцов к матрицам X_r, Y_r. Точнее, после присоединения нового (пусть q-ого) столбца к матрице Y_0 последовательно вычисляются q-ые столбцы матриц X_1, Y_1, X_2, Y_2, и т.д. Добавление столбцов следует прекратить, как только погрешность (2.6) после достаточного числа итераций N станет меньше заданного значения d. Таким образом, указанная модификация процесса (2.5) позволяет определять число k из условия достижения заданной точности разложения. Запишем этот алгоритм в удобном для машинной реализации виде. Введем обозначения:

$$Z \equiv [z_1 \ldots z_m] \in M_{n,m},$$
$$X_q^r \equiv [x_1^r \ldots x_q^r] \in M_{n,q},$$
$$P_q^r \equiv P_{X_q^r},\ P_o^r = E,\ r = 1,\ldots,N,$$
$$Y_q^r \equiv [y_1^r \ldots y_q^r] \in M_{m,q},\ [h_1 \ldots h_q] \in M_{m,q},\ r = 0,1,\ldots,N,\ q \geq 1.$$

А л г о р и т м .

ВХОД: $Z \in M_{n,m}$, точность $d \geq 0$, число итераций $N > 0$.

ВЫХОД: (X_q^N, Y_q^N): $\|Z - X_q^N (Y_q^N)^T\| \leq d$, X_q^N ортонормирована.

1. $H_0 := \|Z\|^2$;
2. $h_{j0} := \|z_j\|^2$ для всех $j = 1,\ldots,m$;
3. ДЛЯ $q = 1,2,\ldots$ ПОВТОРЯТЬ 4-10
4. $\quad y_q^0 := h_q$;
5. $\quad$ДЛЯ $r = 1,\ldots,N$ ПОВТОРЯТЬ 6-7
6. $\quad\quad x_q^r := P_{q-1}^r Z y_q^{r-1} / \|P_{q-1}^r Z y_q^{r-1}\|$;
7. $\quad\quad y_q^r := Z^T x_q^r$;
8. $\quad h_{jq} := h_{j\,q-1} - (y_{jq}^N)^2$ для всех $j = 1,\ldots,m$;
9. $\quad H_q := H_{q-1} - \|y_q^N\|^2$;
10. ПОКА $H_q > d^2$;

Данный алгоритм не гарантирует, что найденная пара матриц (X_q^N, Y_q^N) будет q,d-разложением Z, так как в общем случае для числа столбцов q может не выполняться условие минимальности. Однако, как следует из следующего утверждения, при подходящем выборе параметров алгоритма N и d удается определить одно из k,d-разложений матрицы Z.

У т в е р ж д е н и е . *Пусть выполнены все условия теоремы 3 и* $\mu < 1$. *Тогда для любого* $\varepsilon \in (0, \lambda_k)$ *существует натуральное число* N_0 *такое, что при* $d^2 = \|Z - XY^T\|^2 + \varepsilon$ *и* $N \geq N_0$ *вычисленная алгоритмом пара матриц будет* *k,d-разложением Z.*

Д о к а з а т е л ь с т в о . Легко проверить (см. шаги 1 и 9), что $H_q \equiv \|Z - X_q^N (Y_q^N)^T\|^2$, $q = 1,2,\ldots$ Следовательно условие останова алгоритма (шаг 10) не может выполниться при $q < k$, так как иначе, согласно теореме 2, ε должно было бы превышать λ_k. С другой стороны, выражая номер итерации *r* из оценки (2.6), получим, что при $q = k$ и

$$N \geq 1 + (\ln\mu)^{-1} \ln\left(\frac{1}{tg\varphi\sqrt{k}}\left(\sqrt{1+\varepsilon/d^2}-1\right)\right),$$

условие останова будет выполнено, и вычисленная к этому моменту пара матриц (X_q^N, Y_q^N) будет *k*,*d*-разложением *Z*. Доказано.

Таким образом, при подходящем выборе числа итераций *N* и порога *d* возможно определить одно из *k*,*d*-разложений *Z*. На практике оказалось, что если задавать $N = 2 \div 4$ и $d = (1.2 \div 1.6)\delta$, то за очень редким исключением матрицы X_q^N, Y_q^N не будут содержать лишних столбцов.

Отметим, что при $k << \min(n, m)$ описанный алгоритм позволяет эффективно сжать исходные данные путем замены матрицы *Z* парой матриц существенно меньших размеров. При ограниченных ресурсах памяти это дает возможность хранить обучающие выборки больших объемов.

§3. Оптимальные *k*-разложения

Для обеспечения единственности искомого описания допустимой поверхности необходимо решать задачу регуляризации, накладывая дополнительные ограничения на множество непроизводных поверхностей. Привлекаемая для этого априорная информация должна удовлетворять двум требованиям. Ее использование должно быть обоснованным в рамках решаемой задачи и ее должно быть достаточно, чтобы из бесконечного множества решений выделить единстенное.

Рассмотрим следующий способ учета априорной информации. Будем относить к классу непроизводных поверхностей не все функции вида $f(x)g(y)$, а лишь те, для которых $f(x)$ принадлежит заданному семейству функций $\mathcal{F}$. Функция $g(y)$ по-прежнему может быть любой.

При фиксированном наборе *n* чисел $\xi_i \in \Omega_x$, $i = 1, \ldots, n$ семейство $\mathcal{F}$ индуцирует множество вектор-столбцов

$$\mathcal{F}_n = \left\{[f(\xi_1) \ldots f(\xi_n)]^T : f \in \mathcal{F}\right\}$$

и множество матриц размера $n \times k$ со столбцами из $\mathcal{F}_n$

$$\mathcal{F}_{n,k} = \left\{[u_1 \ldots u_k] : u_s \in \mathcal{F}_n,\ s = 1, \ldots, k\right\}.$$

Пусть (X, Y) — известное *k*-разложение матрицы *Z*, и пусть для простоты столбцы *X* ортонормированы, $X^T X = E$. Задача состоит в том, чтобы выяснить, каким условиям должно удовлетворять семейство $\mathcal{F}$, чтобы существовала единственная (с точностью до перестановки столбцов) матрица $U \in \mathcal{F}_{n,k}$ такая, что $\mathcal{L}(X) = \mathcal{L}(U)$, а также найти способ вычисления этой матрицы.

О п р е д е л е н и е . Семейство функций $\mathcal{F}$ называется *k-нелинейным* по набору точек $(\xi_1 \ldots \xi_n)$, если для любых попарно различных вектор-столбцов $u_0, u_1, \ldots, u_k$ из множества $\mathcal{F}_n$ и любых действительных $a_1, \ldots, a_k$ выполняется условие $a_1 u_1 + \ldots + a_k u_k \neq u_0$.

Семейства функций, обладающие указанным свойством, существуют. Например, параметрическое семейство $\mathscr{F}^{\alpha} = \left\{ e^{-(x-\alpha)^2} : \alpha \in \mathrm{R} \right\}$ k-нелинейно по любому набору из n различных точек для всех $k = 1, \ldots, n-1$. Для доказательства составим $n \times (k+1)$ - матрицу из произвольных попарно различных вектор-столбцов $u_0, u_1, \ldots, u_k$, принадлежащих множеству $\mathscr{F}_n^{\alpha}$. В силу тождества $e^{-(x-\alpha)^2} = e^{-x^2} e^{2x\alpha} e^{-\alpha^2}$ первые ее k+1 строк образуют квадратную матрицу, которая может быть получена умножением строк и столбцов обобщенной матрицы Вандермонда [3] на положительные числа. Определитель последней отличен от нуля. Следовательно вектор-столбцы $u_0, u_1, \ldots, u_k$ линейно независимы, а указанное семейство, в силу их произвольности, k-нелинейно.

Существуют также семейства, не обладающие указанным свойством. В частности, всякое линейное семейство функций, например множество полиномов заданной степени, не может быть k-нелинейным. Как станет ясно ниже, такие семейства функций не могут быть использованы для регуляризации.

О п р е д е л е н и е . Семейство функций $\mathscr{F}$ называется *k-точным* по набору точек $(\xi_1 \ldots \xi_n)$ для данной матрицы $X \in M_{n,k}$ ранга k, если существует невырожденная $k \times k$-матрица A такая, что $XA \in \mathscr{F}_{n,k}$..

Т е о р е м а 4 . *Если $\mathscr{F}$ — семейство функций, k-нелинейное и k-точное по набору точек $(\xi_1 \ldots \xi_n)$ для данной матрицы X, то существует и единственна (с точностью до перестановки столбцов) матрица $U_* \in \mathscr{F}_{n,k}$ такая, что $\mathscr{L}(U_*) = \mathscr{L}(X)$.*

Д о к а з а т е л ь с т в о . В силу k-точности семейства **F** в подпространстве $\mathscr{L}(X)$ найдутся k линейно независимых вектор-столбцов $u_1^*, \ldots, u_k^*$, принадлежащих $\mathscr{F}_n$. В силу k-нелинейности других элементов $\mathscr{F}_n$ в $\mathscr{L}(X)$ быть не может. Следовательно те и только те матрицы $U \in \mathscr{F}_{n,k}$ удовлетворяют условию $\mathscr{L}(X) = \mathscr{L}(U)$, которые составлены из вектор-столбцов $u_1^*, \ldots, u_k^*$, взятых в произвольном порядке. Теорема доказана.

Из доказательства ясно, что k-нелинейность обеспечивает единственность, а k-точность — существование регуляризованного решения.

Заметим, что на практике гарантировать точное совпадение подпространств $\mathscr{L}(X) = \mathscr{L}(U)$ невозможно по многим причинам: из-за погрешностей исходных данных, погрешностей вычисления матрицы X и, самое существенное, неточности описания непроизводных поверхностей элементами семейства. Ослабим поэтому требование k-точности: будем искать матрицу $U_* \in \mathscr{F}_{n,k}$, доставляющую минимум функционалу

$$\bar{J}(U) = \left\| P_U X \right\|^2,$$

который обращается в нуль тогда и только тогда, когда $\mathscr{L}(X) = \mathscr{L}(U)$. Таким образом, для k-точных семейств его минимизация должна приводить к искомому решению.

Ортонормируем столбцы матрицы $U \equiv [u_1 \dots u_k]$, применив к ним процесс Грама-Шмидта, и обозначим полученную матрицу $\tilde{U} \equiv [\tilde{u}_1 \dots \tilde{u}_k]$. Определим последовательность проекционных матриц

$$P_1 = E_n; \quad P_s = P_{\tilde{u}_1} \dots P_{\tilde{u}_{s-1}} \equiv E - \tilde{u}_1 \tilde{u}_1^T - \dots - \tilde{u}_{s-1} \tilde{u}_{s-1}^T; \; s = 2, \dots, k.$$

Тогда

$$\tilde{u}_s = \frac{P_s u_s}{\|P_s u_s\|}, \; s = 1, \dots, k$$

Вследствие элементарных свойств евклидовой матричной нормы

$$\bar{J}(U) = \left\|P_X \tilde{U}\right\|^2 = \sum_{s=1}^{k} \frac{\|P_X P_s u_s\|^2}{\|P_s u_s\|^2}.$$

Заметим, что при наличии k-точности $P_X P_s = P_X$, поэтому функционал $\bar{J}(U)$ можно заменить оценкой сверху

$$J(U) = \sum_{s=1}^{k} \frac{\|P_X u_s\|^2}{\|P_s u_s\|^2}, \tag{3.1}$$

не нарушив при этом требования $J(U) = 0 \Leftrightarrow \mathscr{L}(X) = \mathscr{L}(U)$. Введем обозначения для слагаемых в (3.1)

$$J_s(u) \equiv J_s(u_1, \dots, u_{s-1}, u) \equiv \frac{\|P_X u\|^2}{\|P_s u\|^2}, \; s = 1, \dots, k \tag{3.2}$$

и рассмотрим последовательность задач минимизации

$$u_s^* = \arg \min_{u \in \mathbf{F}_n} J_s(u_1^*, \dots, u_{s-1}^*, u), \; s = 1, \dots, k. \tag{3.3}$$

Если семейство $\mathbf{F}$ удовлетворяет условиям теоремы 4, то последовательная минимизация (3.3) приводит к искомому минимуму функционала $J(U)$. Действительно, в силу k-точности существует набор различных векторов $u_1^*, \dots, u_k^*$, доставляющих нулевые значения функционалам $J_1, \dots, J_k$ соответственно. В силу k-нелинейности этот набор единственен, и входящие в него векторы линейно независимы. Составив из этих векторов матрицу U_*, получим единственную (с точностью до перестановки столбцов) матрицу, доставляющую функционалу $J(U)$ минимум, равный нулю.

Таким образом, минимизацию $J(U)$ по множеству $\mathscr{F}_{n,k}$ удается свести к существенно более простой задаче последовательной минимизации функционалов $J_1, \dots, J_k$ по множеству $\mathscr{F}_n$.

В общем случае, когда минимум $J(U)$ не равен нулю (нет k-точности), можно говорить лишь о его приближеной минимизации с помощью (3.3). Ясно однако, что чем лучше семейство $\mathscr{F}$ описывает функции $f(x)$, тем ближе к нулю минимум функционала $J(U)$, и тем ближе вычисленная с помощью (3.3) матрица U_* к матрице, минимизирующей $J(U)$. Независимо от того, выполняется условие k-точности или нет, можно ввести понятие оптимального k-разложения.

О п р е д е л е н и е. Назовем оптимальным k-разложением пару матриц $(X_0, Y_0) \in M_Z^k$, доставляющую минимум функционалу

$$\Phi(\tilde{X},\tilde{Y}) = \left\| \tilde{X} - U_* \right\|, \ (\tilde{X},\tilde{Y}) \in M_Z^k,$$

где U_* — матрица, вычисленная алгоритмом (3.1-3).

Минимизируя Ф по множеству M_Z^k и используя лемму 2, легко получить формулы для расчета оптимального k-разложения:

$$X_0 = XX^T U_*, \ Y_0 = Y(U_*^T X)^{-1}.$$

Рассмотрим два частных случая.

1. В задачах обучения распознаванию образов [1-2] в качестве семейства $\mathcal{F}$ можно задать конечный набор из l функций $F_1(x),\ldots,F_l(x)$, образующих непроизводные элементы допустимых поверхностей объектов обучающей выборки. В этом случае минимизация каждого из функционалов J_s в (3.2) сводится к линейному перебору l элементов множества $\mathcal{F}$.

Такой подход подразумевает, что мы принимаем гипотезу: «функция $f(x)$ непроизводного элемента допустимой поверхности классифицируемого объекта равна или по меньшей мере очень близка к одной из функций $F_1(x),\ldots,F_l(x)$». Ясно, что решение задачи регуляризации на основе одной лишь обучающей выборки может потребовать увеличения ее длины, неоправданного с точки зрения самого обучения.

2. Пусть $\mathcal{F} = \left\{ f(x,\alpha) : \alpha \in \mathbb{R}^p \right\}$ — параметрическое семейство функций, удовлетворяющее условиям теоремы 4. Минимизация функционалов $J_1,\ldots,J_k$ в (3.3) сводится, очевидно, к минимизации k функций p переменных по множеству допустимых значений параметров. Для решения данной задачи можно воспользоваться известными методами оптимизации [5], причем в качестве начальных приближений удобно брать аппроксимации функций $F(x)$ непроизводных элементов, отобранных из обучающей выборки. Выделить для этой цели k наиболее подходящих объектов обучения можно, применив тот же алгоритм (3.1-3), как это описано в предыдущем случае.

Таким образом, сужение множества непроизводных поверхностей путем априорного задания семейства функций, отвечающего специфическим требованиям точности и нелинейности, и последующее вычисление оптимального k-разложения позволяют выделить единственное структурное описание исходной допустимой поверхности и тем самым решить поставленную задачу.

Работа выполнялась при поддержке Российского Фонда Фундаментальных Исследований (проект 93-012-457).

Список литературы.

[1] Журавлев Ю.И. Об алгебрическом подходе к решению задач распознавания или классификации. // Проблемы кибернетики. 1979. Вып.33. С.5-68.

[2] Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. I-III. / Кибернетика. 1977. 4. С.14-21. 1977. 6. С.21-27. 1978. 2. С.35-43.

[3] Гантмахер Ф.Р. Теория матриц. М.: Наука, 1988.

[4] Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. М.: Наука, 1986.

[5] Васильев Ф.П. Численные методы решения экстремальных задач. М.: Наука, 1988.