

УДК: 004.9

О разложении матриц при помощи метода стохастического градиентного спуска в приложении к задаче направляемой классификации микрочипов

В. Н. Никулин

Вятский государственный университет, ф-т экономики и менеджмента, кафедра ММЭ,
Россия, 610000, г. Киров, ул. Московская, д. 36, Тел.: (8332) 64–48–16

E-mail: vnikulin.uq@gmail.com

Получено 18 марта 2013 г.,
после доработки 5 апреля 2013 г.

Многомерные данные, при использовании значительно большего количества признаков относительно меньшего числа наблюдений, порождают хорошо известную проблему переопределённой задачи. В связи с этим, представляется целесообразным описание данных в терминах меньшего числа мета-признаков, которые вычисляются при помощи так называемых матричных факторизаций. Такие факторизации способствуют уменьшению случайного шума при сохранении наиболее существенной информации. Три новых и взаимосвязанных метода предложены в этой статье: 1) факторизационный механизм градиентного спуска с двумя (согласно размерности микрочипа) гибкими и адаптируемыми параметрами обучения, включая явные формулы их автоматического пересчета, 2) непараметрический критерий для отбора количества факторов, и 3) неотрицательная модификация градиентной факторизации, которая не требует дополнительных вычислительных затрат в сравнении с базовой моделью. Мы иллюстрируем эффективность предложенных методов в приложении к задаче направляемой классификации данных в области биоинформатики.

Ключевые слова: матричная факторизация, ненаправляемое обучение, количество факторов, непараметрический критерий, неотрицательность, оставить одного извне, классификация

On the stochastic gradient descent matrix factorization in application to the supervised classification of microarrays

V. N. Nikulin

Vyatka State University, Faculty of economics and management, department of MME,
36 Moskovskaya st., Kirov, 610000, Russia

Abstract. — Microarray datasets are highly dimensional, with a small number of collected samples in comparison to thousands of features. This poses a significant challenge that affects the interpretation, applicability and validation of the analytical results. Matrix factorizations have proven to be a useful method for describing data in terms of a small number of meta-features, which reduces noise, while still capturing the essential features of the data. Three novel and mutually relevant methods are presented in this paper: 1) gradient-based matrix factorization with two adaptive learning rates (in accordance with the number of factor matrices) and their automatic updates; 2) nonparametric criterion for the selection of the number of factors; and 3) nonnegative version of the gradient-based matrix factorization which doesn't require any extra computational costs in difference to the existing methods. We demonstrate effectiveness of the proposed methods to the supervised classification of gene expression data.

Keywords: matrix factorization, unsupervised learning, number of factors, nonnegativity, bioinformatics, leave-one-out, classification

Citation: *Computer Research and Modeling*, 2013, vol. 5, no. 2, pp. 131–140 (Russian).

© 2013 Владимир Николаевич Никулин

Введение

Матричные факторизации играют важную роль в области биоинформатики, которая напрямую связана с многомерными данными. Отметим существенную особенность, характеризующую информацию на микрочипах: количество наблюдений или пациентов относительно мало в сравнении с количествами генов, которые обычно измеряются в тысячах. Со статистической точки зрения, значительно меньшее число наблюдений в сравнении с числом признаков существенно усложняет задачу классификации. Многие стандартные алгоритмы снижают качество работы от переопределенности при анализе многомерных данных. Как правило, только небольшая часть генов имеет прямое отношение к поставленной задаче. В связи с этим, методы отбора признаков являются неотъемлемой составной частью любого эффективного алгоритма для классификации микрочипов.

Другой метод для уменьшения степени переопределенности связан с описанием данных в терминах меньшего числа метагенов, которые представимы как линейные комбинации оригинальных генов. Недавно, методы независимых/главных компонент и неотрицательной матричной факторизации (nonnegative matrix factorisation, *NMF*) приобрели особую популярность как следствие их очевидной полезности [Oja, 2010]. Отметим, что эти методы являются общими и не требуют каких-либо дополнительных ограничений или какой-либо особой информации.

Количество метагенов относительно невелико (в сравнении с количеством оригинальных генов) и они могут быть эффективны при распознавании образов. Очень часто эти метагены оказываются гораздо более полезными при анализе данных. В связи с этим матричные факторизации, как ненаправляемые методы машинного обучения, широко используются для изучения структуры данных.

В качестве примеров эффективных методов факторизации можно упомянуть метод главных компонент и неотрицательную матричную факторизацию [Lee, 1999]. Дополнительно мы разработали новый и очень быстрый метод для матричной факторизации, основанной на градиентном спуске (gradient-based matrix factorisation, *GMF*) [Nikulin, 2009] в рамках структуры стохастического градиентного спуска (stochastic gradient descent framework).

Матричные разложения

Предположим, что (x_j, y_j) , $j = 1, \dots, n$, — выборка, основанная на результатах наблюдений, где $x_j \in \mathbb{R}^p$ обозначает p -мерный вектор признаков и y_j целевая переменная (известная также как метка). Обозначим через $X = \{x_{ij}, i = 1, \dots, p, j = 1, \dots, n\}$ соответствующую матрицу микрочипа.

Как правило, число генов p исчисляется тысячами, в то время как число экспериментов n исчисляется сотнями. Каждая строка матрицы X соответствует определенному фактору или гену (значения генов, соответствующих конкретным пациентам). Наша цель состоит в нахождении меньшего числа $k \ll p$ мета-факторов. Затем мы сможем выразить или аппроксимировать матрицу наблюдений X в терминах этих мета-факторов. Математически последнее соответствует разложению

$$X \sim AB, \quad (1.1)$$

где матрица A мета-экспериментов, имеющая размерности $p \times k$, и матрица B метагенов, имеющая размерности $k \times n$.

Алгоритм A2GMF, включающий два гибких параметра обучения

Оценим качество матричного разложения (1.1) при помощи следующей функции квадратичного риска

$$L(A, B) = \frac{1}{p \cdot n} \sum_{i=1}^p \sum_{j=1}^n E_{ij}^2, \quad (2.1)$$

где $E_{ij} = x_{ij} - \sum_{f=1}^k a_{if} b_{jf}$.

Важно отметить, что функция риска (2.1) включает $k(p+n)$ регулирующих параметров и может быть нестабильна, если мы будем ее минимизировать без учета взаимной зависимости элементов матриц A и B .

В качестве решения этой проблемы и в соответствии с принципами стохастического градиентного спуска, мы рассмотрим последовательно разности E_{ij} , минимизируя их как функции параметров, которые использованы при определении E_{ij} . В сравнении с обычной градиентной оптимизацией, мы имеем дело в этом случае с двумя взаимно зависимыми множествами параметров и должны последовательно чередовать пересчеты этих параметров.

Следующие частные производные необходимы для алгоритма 1:

$$g_{if} = \frac{\partial E_{ij}^2}{\partial a_{if}} = -2E_{ij} b_{jf},$$

$$h_{jf} = \frac{\partial E_{ij}^2}{\partial b_{jf}} = -2E_{ij} a_{if}.$$

Рассматривая шаг 11 алгоритма 1, мы можем подставить в (2.1) вместо a_{if} новое значение $a_{if} - \lambda_a g_{if}$. Как следствие, (2.1) приобретает следующий вид

$$L(A, B, \lambda_a) = \sum_{i=1}^p \sum_{j=1}^n (E_{ij} + \lambda_a U_{ij})^2, \quad (2.2)$$

где

$$U_{ij} = \sum_{f=1}^k g_{if} \cdot b_{jf} = -2E_{ij} \sum_{f=1}^k b_{jf}^2.$$

Минимизируя (2.2) как функцию λ_a , мы найдем

$$\begin{aligned} \lambda_a &= -\frac{\sum_{i=1}^p \sum_{j=1}^n E_{ij} U_{ij}}{\sum_{i=1}^p \sum_{j=1}^n U_{ij}^2} = \\ &= \frac{1}{2} \frac{\sum_{j=1}^n \sum_{f=1}^k b_{jf}^2 \sum_{i=1}^p E_{ij}^2}{\sum_{j=1}^n \left(\sum_{f=1}^k b_{jf}^2 \right) \sum_{i=1}^p E_{ij}^2} = \frac{1}{2} \frac{\sum_{j=1}^n b_j \phi_j}{\sum_{j=1}^n b_j^2 \phi_j}, \end{aligned} \quad (2.3)$$

где

$$b_j = \sum_{f=1}^k b_{jf}^2, \phi_j = \sum_{i=1}^p E_{ij}^2.$$

Рассматривая шаг 14 алгоритма 1, мы можем подставить в (2.1) вместо b_{fj} новое значение $b_{fj} - \lambda_b h_{fj}$. Как следствие, (2.1) приобретает следующий вид

Алгоритм 1 A2GMF (метод стохастического градиентного спуска).

- 1: Входные данные: X — микрочип.
- 2: Выбор параметров: m — количество глобальных итераций; k — количество факторов; λ_a, λ_b — начальные параметры обучения, $0 < \tau \leq 1$ — корректировочный коэффициент.
- 3: Начальные матрицы A и B генерируются случайным образом.
- 4: Глобальный цикл: начиная с $\ell = 1$ до m повторить шаги 5–16:
- 5: цикл для генов: начиная с $i = 1$ до p повторить шаги 6–15:
- 6: цикл для пациентов: начиная с $j = 1$ до n повторить шаги 7–15:
- 7: вычислить предсказание: $S = \sum_{f=1}^k a_{fj} b_{fj}$;
- 8: вычислить ошибку предсказания: $E = x_{ij} - S$;
- 9: внутренний цикл для факторов: начиная с $f = 1$ до k повторить шаги 10–15:
- 10: вычислить $\alpha = a_{fj} b_{fj}$;
- 11: градиентный пересчет $a_{fj} \leftarrow a_{fj} - \tau(\ell) \cdot \lambda_a \cdot g_{fj}$;
- 12: $E \leftarrow E + \alpha - a_{fj} b_{fj}$;
- 13: вычислить $\alpha = a_{fj} b_{fj}$;
- 14: градиентный пересчет $b_{fj} \leftarrow b_{fj} - \tau(\ell) \cdot \lambda_b \cdot h_{fj}$;
- 15: $E \leftarrow E + \alpha - a_{fj} b_{fj}$;
- 16: пересчет параметров λ_a и λ_b согласно формулам (2.3) и (2.5);
- 17: Выходные данные: A и B — факторные матрицы метагенов.

$$L(A, B, \lambda_b) = \sum_{i=1}^p \sum_{j=1}^n (E_{ij} + \lambda_b V_{ij})^2, \quad (2.4)$$

где

$$V_{ij} = \sum_{f=1}^k h_{fj} \cdot a_{fj} = -2E_{ij} \sum_{f=1}^k a_{fj}^2.$$

Минимизируя (2.4) как функцию λ_b , мы найдем

$$\begin{aligned} \lambda_b &= -\frac{\sum_{i=1}^p \sum_{j=1}^n E_{ij} V_{ij}}{\sum_{i=1}^p \sum_{j=1}^n V_{ij}^2} = \\ &= \frac{1}{2} \frac{\sum_{i=1}^p \sum_{f=1}^k a_{fj}^2 \sum_{j=1}^n E_{ij}^2}{\sum_{i=1}^p \left(\sum_{f=1}^k a_{fj}^2 \right) \sum_{j=1}^n E_{ij}^2} = \frac{1}{2} \frac{\sum_{i=1}^p a_i \psi_i}{\sum_{i=1}^p a_i^2 \psi_i}, \end{aligned} \quad (2.5)$$

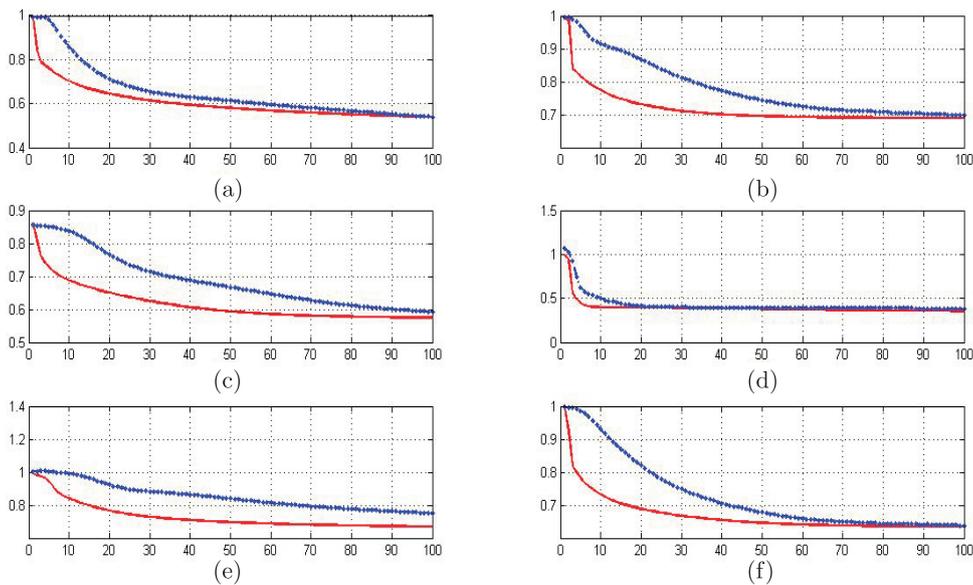


Рис. 1. Поведение риска (2.1) в случае алгоритма *GMF* (синяя штрих-пунктирная линия) и алгоритма 1 (*A2GMF*, красная сплошная линия) (цветное изображение можно увидеть в электронной версии на сайте журнала <http://crm.ics.org.ru/journal/article/2006>)

где

$$a_i = \sum_{j=1}^k a_{ij}^2, \psi_i = \sum_{j=1}^n E_{ij}^2.$$

Рисунок 1 иллюстрирует очевидное превосходство алгоритма *A2GMF* в сравнении с алгоритмом *GMF*. Отметим, что мы использовали фиксированные параметры: $\tau = 1$, $\lambda_a = \lambda_b = 0.01$ в случае алгоритма *GMF* и $\tau = 0.04$ в случае алгоритма *A2GMF*.

Неотрицательная модификация *GMF* (*NN-GMF*)

Во многих случаях реальные данные являются неотрицательными и соответствующие скрытые компоненты/факторы могут найти удовлетворительное объяснение, только если они также неотрицательные.

Предположим, что все элементы матрицы микрочипа X являются неотрицательными. В связи с этим представляется логичным предположить, что факторные матрицы A и B также являются неотрицательными.

Определение *NN-GMF* базируется на очень простой идее. Рассмотрим обобщение (1.1)

$$X \sim \xi(A)\xi(B), \quad (3.1)$$

где ξ — неотрицательная и дифференцируемая функция, матрица $\{\xi(a_{ij}), i = 1, \dots, p, j = 1, \dots, k\}$ имеет такие же размерности, что и матрица A . Формула (3.1) представляет собой гибкую структуру, в которой мы можем использовать любую подходящую функцию. Например, мы можем применить в качестве функции ξ экспоненциальную или квадратичную функции.

Замечание. По определению, условие неотрицательности не требует выполнения каких-либо специальных операций и является естественным следствием неотрицательности преобразующей функции ξ . Таким образом, алгоритм *NN-GMF* не уступает в скорости алгоритму *GMF*.

Согласно [Gao, 2005], тот факт, что задача уменьшения размерности до уровня меньшего, чем количество наблюдений представляется вполне обоснованной. Метод главных компонент способен дать очень хорошие результаты классификации. Однако, осмысление и интерпрета-

ция этих результатов представляется делом затруднительным. Специально для решения этой важной проблемы в статье [Lee, 1999] продемонстрировано, что алгоритм *NMF* производит неотрицательные мета-признаки в области анализа текстов и числовых фотографий с достаточно убедительным истолкованием.

Разреженность и регуляризация

В приложениях разреженные числовые данные является явлением довольно распространенным. Такие данные требуют специального подхода, рассмотрения и изучения, чтобы оптимизировать скорость счета и процесс распределения памяти.

В машинном обучении качество неотрицательности, имеющее интерпретацию как вероятностное распределение, способствует возникновению и усилению свойства разреженности, которое также имеет прямое отношение к отбору признаков.

В связи с этим, разреженность представляется позитивным фактором и для ее достижения мы можем использовать квадратичную регуляризацию, которая штрафует за чрезмерное использование элементов матриц A и B , принимая во внимание их количество и качество (абсолютные величины значений).

Случай, когда значительная часть данных утрачена или отсутствует

Предположим, что некоторые элементы матрицы X отсутствуют. Поскольку в процессе любой глобальной итерации мы рассматриваем не все элементы сразу, а последовательность элементов, мы можем исключить из этой последовательности любые отобранные элементы. Таким образом, мы можем рассматривать, используя алгоритмы в рамках структуры *GMF*, матрицы огромных размеров. Например, в случае широко известного Netflix Cup (оценка вкусов и предпочтений клиентов в области кинопроката) мы имеем дело с сотнями тысяч клиентов и десятками тысяч фильмов, в то время как только около 5 % от общего количества возможных отношений известно.

Непараметрический и ненаправляемый (без учителя) критерий для определения числа факторов

Основная идея непараметрической статистики состоит в построении оценок при использовании минимальных предположений и допущений. Критерий, который мы предлагаем в этой секции, является наиболее общим и может привлечь внимание в качестве самоцели, но не только как прикладной метод для определения числа мета-переменных. Естественно предположить, что матрица микрочипа содержит некоторые системные зависимости, которые будут обнаружены при использовании факторизационного метода. Мы можем легко разрушить эти зависимости, используя случайные перестановки всех колонок матрицы, рассматривая их как векторы. Заметим, что перестановки отдельных колонок осуществляются независимо.

Мы обозначим через $X^{(\gamma)}$, $\gamma = 1, 2, 3$, три случайные перестановки матрицы X . Через $A^{(\gamma)}$ и $B^{(\gamma)}$ мы обозначим соответствующие факторные матрицы. Рисунок 2 иллюстрирует (черная сплошная линия) поведение

$$D_k = \frac{1}{3} \sum_{\gamma=1}^3 \Phi_k^{(\gamma)} - \Phi_k, \quad (4.1)$$

как функцию k , где

$$\Phi_k = \sqrt{\frac{L(A_k, B_k)}{pn}}, \Phi_k^{(\gamma)} = \sqrt{\frac{L(A_k^{(\gamma)}, B_k^{(\gamma)})}{pn}}, \gamma = 1, 2, 3.$$

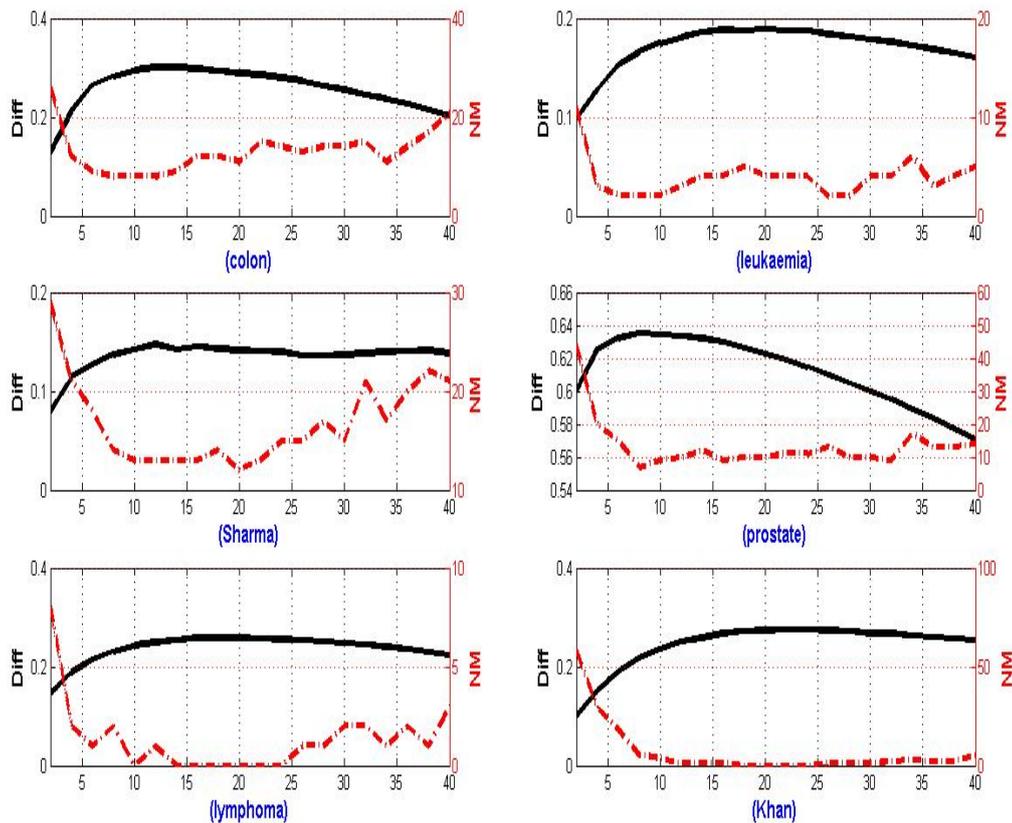


Рис. 2. Непараметрический критерий для определения числа факторов, где левая шкала соответствует (4.1) — черная сплошная линия. Правая шкала соответствует числу неверно классифицированных наблюдений как функция k (красная штрих-пунктирная линия), где мы использовали оценивающую схему $A2GMF+LOO\{lm\}$ применительно к каждой базе данных (цветное изображение можно увидеть в электронной версии на сайте журнала <http://crm.ics.org.ru/journal/article/2006>)

В соответствии с определением (4.1), значения D_k всегда положительны. Отметим, что значения D_k невелики, если количество факторов мало. Этот факт имеет очень простое истолкование: малого числа факторов недостаточно, чтобы охватить скрытые зависимости и отношения, содержащиеся в микрочипе. Затем значения D_k будут расти до некоторого уровня. Предлагается использовать высшую точку графика в качестве критерия для определения числа мета-переменных. После этой точки значения D_k будут уменьшаться, как следствие переопределённости.

Данные для экспериментов

Мы выбрали для наших экспериментов шесть широко известных реальных баз данных (colon, leukaemia, Sharma, Khan, lymphoma, prostate), которые могут быть получены при помощи интернета [Statnikov, 2008]. Все базы данных включают матрицу X (микрочип), в которой каждая из p строк представляет значения отдельного гена для различных пациентов. Кроме того, имеются значения целевой функции, которые разбивают множество пациентов на пересекающиеся классы. В четырех случаях из шести мы имеем дело с двоичной или бинарной задачей (binary, два класса), в оставшихся двух случаях (Khan, lymphoma) количество классов превышает два (multi-class). Отметим, что двойная нормализация была применена ко всем базам данных. Во-первых, мы использовали линейное преобразование применительно к каждому столбцу с тем, чтобы соответствующее среднее было равно нулю а стандартное отклонение

было равно единице. Затем, мы применили аналогичное преобразование к каждой строке микрочипа.

Таблица 1. Классификационные результаты (количества неверно классифицированных наблюдений) в случае схемы: $A2GMF + LOO\{lm\}$

k	colon	leukaemia	Sharma	Khan	lymphoma	prostate
2	22	8	24	48	6	44
4	10	5	22	27	2	20
6	9	2	18	15	1	17
8	7	1	14	5	0	7
10	8	1	9	6	0	9
12	8	1	10	2	0	10
14	8	2	9	0	0	12
16	10	2	10	0	0	9
18	9	2	10	0	0	10
20	10	2	11	0	0	10
22	11	5	10	0	0	11
24	13	3	13	0	0	11
26	13	1	15	1	1	13
28	11	1	16	1	1	10
30	14	2	16	1	2	10
32	13	3	16	3	1	9
34	9	2	18	2	1	17
36	12	3	18	3	1	13
38	14	4	18	3	2	13
40	15	6	20	3	3	14

Результаты экспериментов

После разложения оригинальной матрицы X мы использовали классическую оценивающую схему «убрать одного во-вне» (leave-one-out, LOO) применительно к матрице метагенов B . Это означает, что мы изъяли из рассмотрения наблюдение с некоторым индексом i и использовали оставшиеся $n - 1$ наблюдения для формирования классификатора, который затем был использован для предсказания метки соответствующей изъятому наблюдению i . В качестве базового классификатора мы использовали функцию lm в R , которая соответствует стандартной регрессионной модели без каких-либо регулирующих параметров.

Процедура для оценки качества классификатора

Мы обозначим [Nikulin, 2009] описанную выше процедуру как

$$A2GMF + LOO\{lm\}. \quad (5.1)$$

Обозначение (5.1) имеет естественную интерпретацию как последовательность двух шагов: 1) вычислить матрицу метагенов B используя алгоритм 1 (уменьшение размерности), и затем 2) использовать оценивающую схему LOO с функцией lm в качестве базового классификатора.

Замечание. Поскольку алгоритм 1 является ненаправляемым (без учителя или без использования меток), нет необходимости включать этот алгоритм во внутренний цикл схемы (5.1). Этот факт очень существенен и позволяет значительно снизить вычислительные затраты.

Результаты классификации

Как отмечено в [Nikulin, 2009], число факторов/метагенов не должно быть слишком большим (чтобы уменьшить эффект перетренировки) и не должно быть слишком маленьким. В последнем случае модель будет страдать от чрезмерного сглаживания и, как следствие, от потери существенной информации, см. рисунок 2 и таблицу 1.

Таблица 2. Сравнение рекомендованных чисел факторов (метагенов) k_{NP} и k^* , где k_{NP} был определен согласно непараметрическому критерию как точка на горизонтальной оси, которая соответствует максимуму (4.1), см. рисунок 2а–е. Подходящий интервал значений для k^* был определен согласно схеме $A2GMF + LOO\{lm\}$; столбец «NM» показывает числа неверно классифицированных наблюдений

Data	n	p	k^*	k_{NP}	NM
colon	62	2000	8–14	12	7
leukaemia	72	1896	8–20	18	1
Sharma	60	1368	10–22	18	9
khan	83	2308	14–30	22	0
lymphoma	62	4026	8–28	18	0
prostate	102	6033	8	8	7

В наших предыдущих работах [Nikulin, 2009] и [Nikulin, 2011] мы представили результаты классификации, основанные на модели $GMF + LOO\{lm\}$. Эти результаты не уступают соответствующим результатам, опубликованным в статьях [Dettling, 2003], [Hennig, 2007] и [Statnikov, 2008].

Отметим, что результаты таблиц 1 и 2 могут быть улучшены, если мы применим вместо функции lm более совершенный классификатор при использовании дополнительной настройки регулирующих параметров в зависимости от конкретных данных. Например, мы можем рекомендовать методы опорных векторов и так называемые случайные леса, которые использовались в [Statnikov, 2008].

Количества факторов k_{NP} и k^* , представленные в таблице 2, были вычислены при использовании совершенно разных методов. Однако, мы можем видеть довольно близкое соответствие между k_{NP} и k^* , где значение k_{NP} было определено согласно непараметрическому критерию как точка на горизонтальной оси, соответствующая максимуму дистанции (4.1), см. рисунок 2; подходящий интервал для k^* был выбран в соответствии со схемой $A2GMF + LOO\{lm\}$. Интересно отметить, что во всех шести случаях (соответственно рассматриваемым базам данных) рекомендации сделанные при помощи непараметрического (без учителя) метода были подтверждены другим направляемым методом (в случае микрочипа с идентификатором «Prostate» мы имеем точное соответствие). Этот факт может служить подтверждением того, что непараметрический метод способен раскрыть некоторые глубинные статистические закономерности, содержащиеся в микрочипах.

Затраты времени

Для вычислений был использован компьютер фирмы Acer, Duo 2.2GHz, 3GB RAM. Для 100 глобальных итераций (см. траектории представленные на рисунке 1) в случае микрочипа Colon потребовалось 6 секунд для GMF и 8 секунд для $A2GMF$ затраты времени для других баз данных были аналогичные).

Заключительные замечания

В этой статье мы представили три новых идеи и продемонстрировали их прикладную значимость. GMF представляет собой гибкую структуру, где мы можем использовать одну или

несколько перспективных моделей в качестве независимых компонентов. В частности, мы можем использовать регуляризацию элементов факторных матриц или бустинг [Dettling, 2003].

Представляется естественным использовать различные параметры обучения в применении к различным факторным матрицам. Кроме того, величины параметров обучения не должны быть фиксированным и предлагается обновлять их после каждой глобальной итерации согласно данным формулам. Как показал наш эксперимент, алгоритм *A2GMF*, представленный в этой статье, значительно быстрее. Это означает, что меньшее количество глобальных итераций требуется для того, чтобы обеспечить такое же качество матричного разложения (в смысле критерия (1.1)).

Предлагаемый непараметрический критерий для выбора количества факторов/метагенов является общим. В качестве альтернативы многочисленным экспериментам с использованием схемы *LOO*, предлагаемый метод позволяет значительно снизить временные затраты. Мы уверены, что в ряде приложений возможно использование аналогичных методов, которые основаны на принципиальных идеях, сформулированных в данной работе.

Неотрицательная версия *GMF* представляет собой третье предлагаемое новшество, и для ее реализации не требуются какие-либо дополнительные расходы времени вычислений. Как следствие, *NN-GMF* не уступает по скорости оригинальной версии *GMF*. Логично ожидать, что специалисты, работающие во многих прикладных областях, найдут этот алгоритм вполне конкурентоспособным в сравнении с хорошо известным алгоритмом *NMF* [Lee, 1999].

Автор благодарен рецензенту за полезные замечания, которые помогли существенно улучшить статью.

Список литературы

- Dettling M., Buhlmann P. Boosting for tumor classification with gene expression data // *Bioinformatics*. — 2003. — Т. 19(9). — P. 1061–1069.
- Gao Y., Church G. Improving molecular cancer class discovery through sparse nonnegative matrix factorization // *Bioinformatics*. — 2005. — Т. 21(21). — P. 3970–3975.
- Hennig C. Cluster-wise assessment of cluster stability // *Computational Statistics and Data Analysis*. — 2008. — Т. 52. — P. 258–271.
- Lee D., Seung H. Learning the parts of objects by nonnegative matrix factorization // *Nature*. — 1999. — Т. 401. — P. 788–791.
- Nikulin V., Huang N.-Y., Ng S.-K., Rathnayake S., McLachlan G. A very fast algorithm for matrix factorization // *Statistics and Probability Letters*. — 2011. — Т. 81. — P. 773–782.
- Nikulin V., McLachlan G. On a general method for matrix factorization applied to supervised classification // *Proceedings of the 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, Washington D.C. — 2009. — P. 44–49.
- Oja E., Ilin A., Luttinen J., Yang Z. Linear expansions with nonlinear cost functions: modelling, representation, and partitioning // *Plenary and Invited Lectures, Barcelona, Spain, WCCI 2010*, Edited by Joan Aranda and Sebastia Xambo. — 2010. — P. 105–123.
- Statnikov A., Wang L., Aliferis C. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification // *BMC Bioinformatics*. — 2008. — Т. 9(319).