

# Метрическое обучение и снижение размерности пространства в задачах кластеризации\*

*Р. В. Исаченко, А. М. Катруца*

*isa-ro@yandex.ru, amkatrutsa@yandex.ru*

Московский физико-технический институт, г. Долгопрудный, Институтский пер., 9

Работа посвящена использованию методов метрического обучения в задачах кластеризации. Применение метрического обучения позволяет модифицировать расстояния между объектами, сближая объекты из одного кластера и отдаляя объекты из разных кластеров. В данной работе расстояние измеряется при помощи метрики Махаланобиса. Процедура метрического обучения состоит в определении оптимальной матрицы ковариаций множества объектов. Кластеризация осуществляется алгоритмом  $k$ -средних и алгоритмом адаптивного метрического обучения, понижающим размерность признакового пространства. Для сравнения этих методов произведен вычислительный эксперимент на синтетических и реальных данных, сделан вывод об эффективности рассматриваемых методов.

**Ключевые слова:** *кластеризация;  $k$ -средних; алгоритм адаптивного метрического обучения; EM-алгоритм*

DOI: 10.21469/22233792.2.1.02

## 1 Введение

Методы метрического обучения [1, 2] применяются при идентификации лиц [3], классификации текстовых документов [4], распознавании рукописных цифр [5]. В данной работе решается задача метрического обучения при кластеризации объектов на заданное количество кластеров [6]. Требуется выявить наборы объектов, похожих между собой, причем степень сходства определяется расстоянием между объектами. Целью метрического обучения является выбор оптимального способа измерения расстояния между объектами.

Для решения данной задачи в работе [7] используются деревья принятия решений [8], основанные на логических схемах. Но проблема построения оптимального дерева является NP-полной, и практическое применение данного метода основано на эвристических алгоритмах [9]. Другой подход к кластеризации используется в статье [10], где предложен эффективный способ отбора признаков, а в качестве алгоритма кластеризации выбран EM (expectation-maximization) алгоритм. В работе [11] для кластеризации объектов применяется метрическое обучение. Эта работа используется в настоящем исследовании в качестве базовой.

Предлагаемый алгоритм адаптивного метрического обучения объединяет задачи кластеризации и метрического обучения в одну задачу максимизации функционала качества. Ключевой идеей алгоритма адаптивного метрического обучения является понижение размерности пространства объектов таким образом, чтобы расстояния между кластерами были максимальны. Для решения оптимизационной задачи используется EM-подход. На каждой итерации алгоритм находит ортогональное преобразование, понижающее размерность признакового пространства, и кластеризует объекты в новом пространстве.

В данной работе алгоритм адаптивного метрического обучения применяется к синтетическим и реальным данным. Цель эксперимента — показать работоспособность предложенного подхода и провести его сравнение с базовым алгоритмом. В качестве базового для

---

\*Проект поддержан грантом РФФИ № 16-37-00485.

сравнения алгоритма выбран алгоритм  $k$ -средних [12]. Данный алгоритм минимизирует суммарное квадратичное отклонение точек кластеров от центров этих кластеров. Получена оценка качества работы построенного алгоритма. Проведен сравнительный анализ результатов, полученных с помощью метрического обучения и без него.

## 2 Постановка задачи метрического обучения

Пусть  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{T \times N}$  — множество объектов. Объект  $\mathbf{x}_i = [x_i^1, \dots, x_i^T]^\top$  задан в виде вектора в пространстве признаков. Требуется выявить кластерную структуру данных и разбить множество объектов  $\mathbf{X}$  на множество непересекающихся кластеров, т. е. построить отображение

$$a : \mathbf{X} \rightarrow \{1, \dots, K\}.$$

Обозначим  $y_i = a(\mathbf{x}_i)$ ,  $y_i \in \{1, \dots, K\}$ , — метка кластера объекта  $\mathbf{x}_i$ . Необходимо выбрать метки кластеров  $\{y_i\}_{i=1}^N$  таким образом, чтобы расстояния между кластерами были максимальными. Центр  $\boldsymbol{\mu}$  множества объектов  $\mathbf{X}$  и центры кластеров  $\{\boldsymbol{\mu}_k\}_{k=1}^K$  вычисляются по формулам:

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i; \quad \boldsymbol{\mu}_k = \frac{\sum_{i=1}^N [y_i = y_k] \mathbf{x}_i}{\sum_{i=1}^N [y_i = y_k]}. \quad (1)$$

Введем на множестве объектов  $\mathbf{X}$  расстояние Махаланобиса

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{A}^{-1} (\mathbf{x}_i - \mathbf{x}_j)}, \quad (2)$$

где  $\mathbf{A}$  — это матрица ковариаций множества  $\mathbf{X}$

$$\mathbf{A} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top. \quad (3)$$

**Определение 1.** Функционалом качества кластеризации  $Q$  назовем межкластерное расстояние:

$$Q(\{\boldsymbol{\mu}_k\}_{k=1}^K) = \sum_{k=1}^K N_k \rho^2(\boldsymbol{\mu}_k, \boldsymbol{\mu}),$$

где  $N_k = \sum_{i=1}^N [y_i = y_k]$  — число объектов в кластере  $k$ .

Поставим задачу кластеризации как задачу максимизации функционала

$$Q(\{\boldsymbol{\mu}_k\}_{k=1}^K) \rightarrow \max_{\boldsymbol{\mu}_k \in \mathbb{R}^T}. \quad (4)$$

Для улучшения качества решения этой задачи предлагается применить метод метрического обучения к ковариационной матрице  $\mathbf{A}$ . Найдем такую матрицу  $\mathbf{A}$ , для которой функционал качества принимает максимальное значение:

$$\mathbf{A}^* = \arg \max_{\mathbf{A} \in \mathbb{R}^{T \times T}} Q(\{\boldsymbol{\mu}_k^*\}_{k=1}^K), \quad (5)$$

где  $\{\boldsymbol{\mu}_k^*\}_{k=1}^K$  — решение задачи кластеризации (4).

## 3 Алгоритм адаптивного метрического обучения

Для решения поставленных оптимизационных задач (4), (5) используется алгоритм адаптивного метрического обучения. Предлагается понизить размерность пространства

объектов  $\mathbf{X}$  с помощью линейного ортогонального преобразования  $\mathbf{G} \in \mathbb{R}^{T \times L}$ ,  $\mathbf{G}^\top \mathbf{G} = \mathbf{I}$ , где новая размерность  $L < T$

$$\mathbf{X} \ni \mathbf{x}_i \mapsto \hat{\mathbf{x}}_i = \mathbf{G}^\top \mathbf{x}_i \in \mathbb{R}^L, \quad i = 1, \dots, N.$$

Центр  $\hat{\boldsymbol{\mu}}$  множества объектов  $\{\hat{\mathbf{x}}_i\}_{i=1}^N$  вычисляется по формуле (1). Расстояния между объектами вычисляются по формуле (2), где в качестве матрицы  $\hat{\mathbf{A}}$  используется матрица ковариаций (3) множества объектов  $\{\hat{\mathbf{x}}_i\}_{i=1}^N$

$$\hat{\mathbf{A}} = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{x}}_i - \hat{\boldsymbol{\mu}})(\hat{\mathbf{x}}_i - \hat{\boldsymbol{\mu}})^\top = \frac{1}{N} \sum_{i=1}^N \mathbf{G}^\top (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top \mathbf{G} = \mathbf{G}^\top \mathbf{A} \mathbf{G}.$$

**Определение 2.** Индикаторной матрицей назовем матрицу  $\mathbf{F} = \{\delta_{ik}\} \in \mathbb{R}^{N \times K}$ , где

$$\delta_{ik} = \begin{cases} 1, & \text{если } a(\mathbf{x}_i) = y_k; \\ 0, & \text{если } a(\mathbf{x}_i) \neq y_k. \end{cases}$$

**Определение 3.** Взвешенной индикаторной матрицей назовем матрицу  $\mathbf{L} = \mathbf{F}(\mathbf{F}^\top \mathbf{F})^{-1/2} = \{l_{ik}\} \in \mathbb{R}^{N \times K}$ , элементы которой равны:

$$l_{ik} = \begin{cases} \frac{1}{\sqrt{N_k}}, & \text{если } a(\mathbf{x}_i) = y_k; \\ 0, & \text{если } a(\mathbf{x}_i) \neq y_k. \end{cases}$$

**Утверждение 1.** С использованием данных обозначений задача кластеризации (4) и задача метрического обучения (5) сводятся к общей задаче максимизации функционала качества [13]

$$Q = \frac{1}{N} \text{tr}(\mathbf{L}^\top \mathbf{X}^\top \mathbf{G} \hat{\mathbf{A}}^{-1} \mathbf{G}^\top \mathbf{X} \mathbf{L}) = \frac{1}{N} \text{tr}(\mathbf{L}^\top \mathbf{X}^\top \mathbf{G} (\mathbf{G}^\top \mathbf{A} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{X} \mathbf{L}) \rightarrow \max_{\mathbf{G}, \mathbf{L}}. \quad (6)$$

## 4 Решение задачи метрического обучения

Для решения задачи (6) алгоритм адаптивного метрического обучения использует EM-подход. На каждом шаге итеративно вычисляются локальные оптимальные значения матриц  $\mathbf{G}$  и  $\mathbf{L}$ . На  $E$ -шаге необходимо найти матрицу  $\mathbf{L}$ , которая является решением оптимизационной задачи (6) при фиксированной матрице  $\mathbf{G}$ . В качестве начального приближения получим взвешенную индикаторную матрицу  $\mathbf{L}$  с помощью алгоритма кластеризации  $k$ -средних с евклидовой метрикой. На  $M$ -шаге производится нахождение оптимального значения матрицы  $\mathbf{G}$  при фиксированной матрице  $\mathbf{L}$ . Алгоритм завершается при стабилизации функционала  $Q$  на последовательности итераций.

### 4.1 Алгоритм $k$ -средних

В данной работе базовым алгоритмом для сравнения является алгоритм  $k$ -средних. Первым шагом алгоритм выбирает из множества  $\mathbf{X}$  случайным образом  $K$  объектов  $\{\boldsymbol{\mu}_k\}_{k=1}^K$  — начальные центры кластеров. Для каждого объекта  $\mathbf{x}_i$  вычисляется расстояние (2) до каждого центра кластера  $\boldsymbol{\mu}_k$  с единичной матрицей трансформаций. Объект  $\mathbf{x}_i$  относится к кластеру, расстояние до которого оказалось наименьшим. Далее производится

вычисление новых центров кластеров по формуле (1). Алгоритм завершается, если значения центров кластеров прекращают меняться.

#### 4.2 Оптимизация матрицы $\mathbf{G}$ с фиксированной матрицей $\mathbf{L}$

Для любых двух квадратных матриц  $\mathbf{A}$  и  $\mathbf{B}$  справедливо  $\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA})$ . Данное свойство позволяет переформулировать задачу (6) следующим образом:

$$Q = \frac{1}{N} \text{tr}(\mathbf{L}^\top \mathbf{X}^\top \mathbf{G} (\mathbf{G}^\top \mathbf{A} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{X} \mathbf{L}) = \frac{1}{N} \text{tr}((\mathbf{G}^\top \mathbf{A} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{X} \mathbf{L} \mathbf{L}^\top \mathbf{X}^\top \mathbf{G}).$$

**Утверждение 2.** Обозначим  $\mathbf{B} = \mathbf{X} \mathbf{L} \mathbf{L}^\top \mathbf{X}^\top$ . Обозначим через  $\mathbf{G} = [\mathbf{v}_1, \dots, \mathbf{v}_K]$  матрицу, состоящую из  $K$  собственных векторов матрицы  $\mathbf{A}^{-1} \mathbf{B}$ , отвечающих наибольшему собственным значениям. Тогда решением (6) является ортогональная матрица, полученная  $QR$ -разложением матрицы  $\mathbf{G}$ .

Функционал качества  $Q$  зависит только от матрицы  $\mathbf{G}$ . Обозначим

$$s(\mathbf{G}) = \text{tr}((\mathbf{G}^\top \mathbf{A} \mathbf{G})^{-1} \mathbf{G}^\top \mathbf{B} \mathbf{G}).$$

На данном шаге задача (6) принимает вид:

$$\mathbf{G}^* = \arg \max_{\mathbf{G} \in \mathbb{R}^{T \times L}} s(\mathbf{G}); \quad (7)$$

$$\mathbf{G}^\top \mathbf{G} = \mathbf{I}. \quad (8)$$

Ранг произведения матриц не превосходит рангов сомножителей, поэтому ранг матрицы  $\mathbf{B}$  не превосходит  $K$ . Решением (7) является матрица  $\mathbf{G} = [\mathbf{v}_1, \dots, \mathbf{v}_K]$ , состоящая из  $K$  собственных векторов матрицы  $\mathbf{A}^{-1} \mathbf{B}$ , отвечающих наибольшему собственным значениям. Таким образом, размерность нового пространства объектов будет равна количеству кластеров  $K$ .

В общем случае матрица  $\mathbf{G}$  не является ортогональной. Заметим, что для любой невырожденной матрицы  $\mathbf{G}$  верно  $s(\mathbf{G}) = s(\mathbf{G} \mathbf{M})$ . Для учета условия ортогональности (8) найдем  $QR$ -разложение матрицы  $\mathbf{G}$ . Тогда ортогональная матрица  $\mathbf{Q}$  является оптимальным значением  $\mathbf{G}^*$ .

#### 4.3 Оптимизация матрицы $\mathbf{L}$ с фиксированной матрицей $\mathbf{G}$

**Утверждение 3.** Обозначим  $\hat{\mathbf{K}} = (1/N) \mathbf{X}^\top \mathbf{G} \hat{\mathbf{A}}^{-1} \mathbf{G}^\top \mathbf{X}$ . Тогда задача (6) эквивалентна задаче кластеризации  $k$ -средних с заданным ядром  $\hat{\mathbf{K}}$  [14].

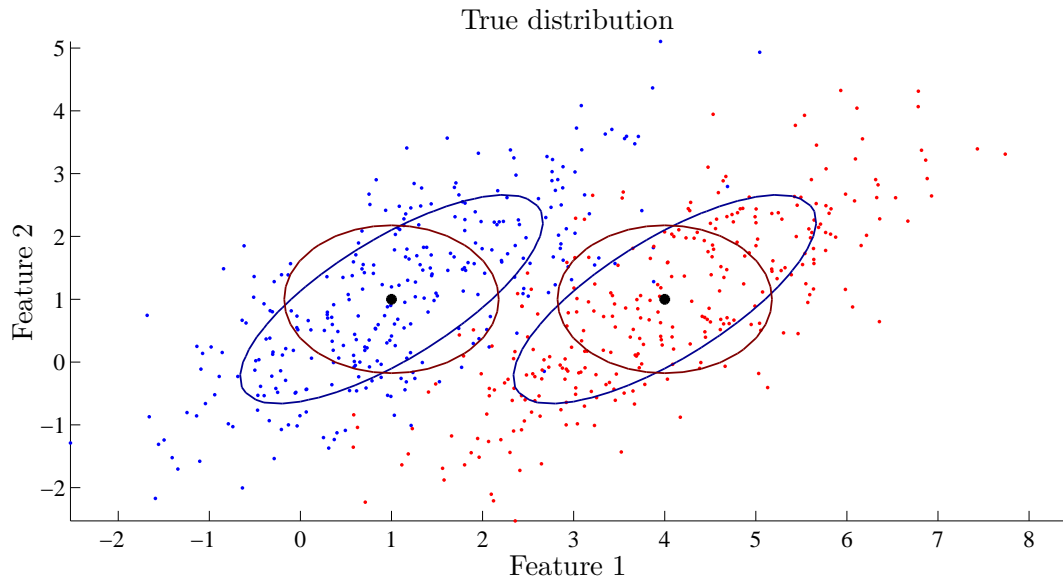
При фиксированной матрице  $\mathbf{G}$  задача (6) принимает вид:

$$\text{tr}(\mathbf{L}^\top \hat{\mathbf{K}} \mathbf{L}) \rightarrow \max_{\mathbf{L} \in \mathbb{R}^{N \times K}}.$$

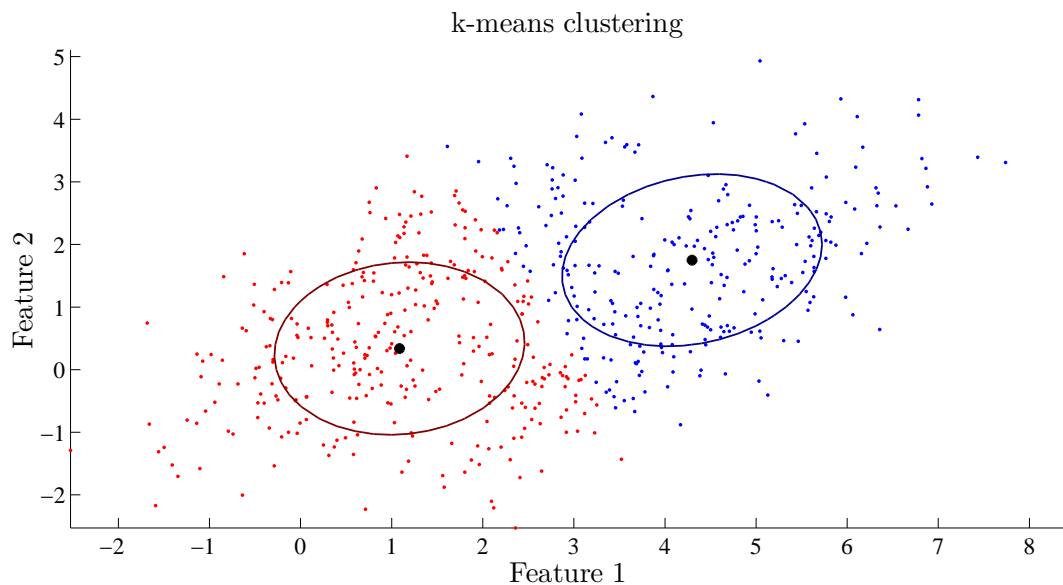
Матрица  $\hat{\mathbf{K}}$  является симметричной и неотрицательно определенной, тем самым может быть выбрана в качестве ядра.

## 5 Вычислительный эксперимент

В целях проверки работоспособности предложенного подхода проведен вычислительный эксперимент на модельных данных. Сгенерирована выборка объектов, принадлежащих одному из двух классов, в двумерном пространстве. Каждый объект принадлежит



**Рис. 1** Истинное распределение двумерных модельных данных



**Рис. 2** Результат кластеризации алгоритмом  $k$ -средних

многомерному нормальному распределению. На рис. 1 показано истинное распределение объектов, черным цветом выделены истинные центры классов и линии уровня функции распределения.

Применим к данной выборке базовый алгоритм  $k$ -средних. Результат кластеризации показан на рис. 2, где черным цветом выделены найденные центры классов и линии уровня функции распределения, построенной по выборочной ковариационной матрице.

Взяв за начальное приближение результаты работы алгоритма  $k$ -средних, проведем кластеризацию с помощью алгоритма адаптивного метрического обучения. Результаты работы алгоритма продемонстрированы на рис. 3.

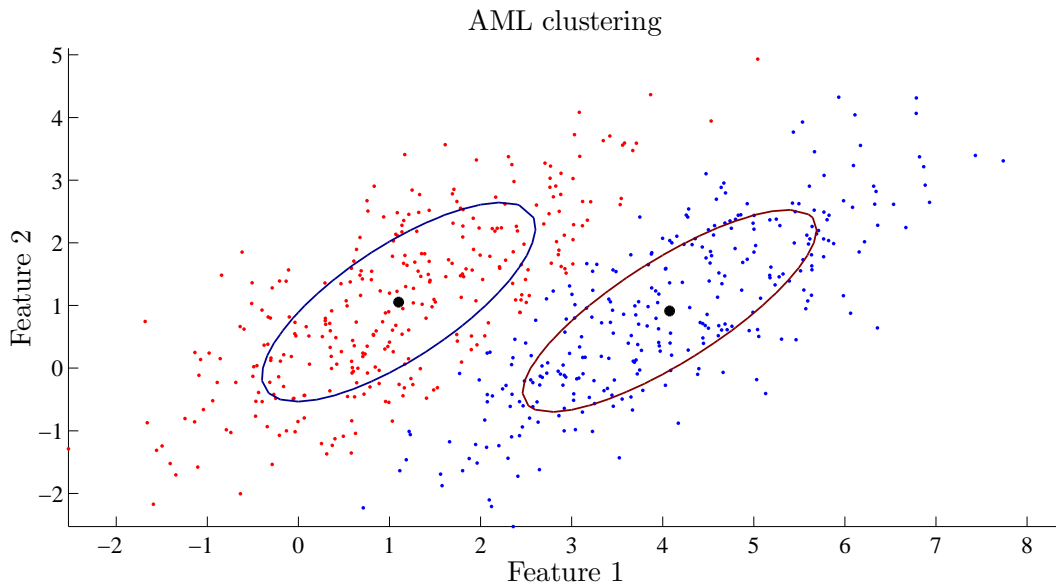


Рис. 3 Результат кластеризации алгоритмом адаптивного метрического обучения

Таблица 1 Результаты кластеризации

Выборка	Качество кластеризации	
	$k$ -средних	AML
Letter Recognition	0,356	<b>0,428</b>
Optical Recognition of Handwritten Digits	0,758	<b>0,790</b>
Seeds	0,833	<b>0,881</b>
Image Segmentation	0,545	<b>0,737</b>
Breast Cancer Wisconsin	<b>0,960</b>	0,956

На рисунках заметно улучшение результатов кластеризации. Измеренная точность кластеризации алгоритма  $k$ -средних составила 0,76, алгоритма адаптивного метрического обучения — 0,94, что говорит о работоспособности данного подхода.

Таблица 1 показывает результаты вычислительного эксперимента на реальных данных. Алгоритм был применен к 5 выборкам, взятых из репозитория UCI [15–19]. Оценкой качества кластеризации служит число правильно кластеризованных объектов. При кластеризации объектов на более чем два класса возникает проблема соотношения истинных классов с полученными кластерами. Данная проблема была формализована в виде задачи о назначениях и решена с помощью венгерского алгоритма. Вычислительный эксперимент на реальных данных показал увеличение точности кластеризации при использовании метрического обучения.

## 6 Заключение

В данной работе предложен новый способ снижения размерности задачи кластеризации объектов на заданное число кластеров. Сравнивались результаты кластеризации базового алгоритма  $k$ -средних и алгоритма адаптивного метрического обучения. Проведен вычислительный эксперимент на синтетических данных. Он показал наглядную интерпретацию алгоритма адаптивного метрического обучения и улучшение качества кластериза-

ции. Вычислительный эксперимент на реальных данных показал эффективность данного подхода в реальных задачах.

Авторы выражают благодарность В. В. Стрижову за постановку задачи и внимательное отношение к работе.

## Литература

- [1] Yang L., Jin R. Distance metric learning: A comprehensive survey. Michigan State University, 2006. Vol. 2. 51 p.
- [2] Bellet A., Habrard A., Sebban M. A survey on metric learning for feature vectors and structured data. ArXiv:1306.6709, 2013.
- [3] Guillaumin M., Verbeek J., Schmid C. Is that you? Metric learning approaches for face identification // IEEE 12th Conference (International) on Computer Vision Proceedings, 2009. P. 498–505.
- [4] Yang L., Jin R., Sukthankar R., Liu Y. An efficient algorithm for local distance metric learning // AAAI, 2006. Vol. 2. P. 543–548.
- [5] Globerson A., Roweis S. T. Metric learning by collapsing classes // Advances in neural information processing systems / Eds. Y. Weiss, B. Schölkopf, J. Platt. — Cambridge, MA, USA: MIT Press, 2005. Vol. 18. P. 451–458.
- [6] Xing E. P., Ng A. Y., Jordan M. I., Russell S. Distance metric learning with application to clustering with side-information // Advances in Neural Information Processing Systems, 2003. Vol. 15. P. 505–512.
- [7] Geurts P. Pattern extraction for time series classification // Principles of data mining and knowledge discovery. — Springer, 2001. P. 115–127.
- [8] Friedl M. A., Brodley C. E. Decision tree classification of land cover from remotely sensed data // Remote Sensing of Environment, 1997. Vol. 61. No. 3. P. 399–409.
- [9] Hyafil L., Rivest R. L. Constructing optimal binary decision trees is np-complete // Inform. Proc. Lett., 1976. Vol. 5. No. 1. P. 15–17.
- [10] Dy J. G., Brodley C. E. Feature selection for unsupervised learning // J. Machine Learning Res., 2004. Vol. 5. P. 845–889.
- [11] Ye J., Zhao Z., Liu H. Adaptive distance metric learning for clustering // IEEE Conference on Computer Vision and Pattern Recognition, 2007. P. 1–7.
- [12] MacQueen J., et al. Some methods for classification and analysis of multivariate observations // 5th Berkeley Symposium on Mathematical Statistics and Probability Proceedings, 1967. Vol. 1. P. 281–297.
- [13] Ding C., He X., Simon H. D. On the equivalence of nonnegative matrix factorization and spectral clustering // SIAM Data Mining Conference Proceedings, 2005. P. 606–610.
- [14] Shawe-Taylor J., Cristianini N. Kernel methods for pattern analysis. — Cambridge University Press, 2004. 478 p.
- [15] Letter recognition dataset. <http://archive.ics.uci.edu/ml/machine-learning-databases/letter-recognition/>.
- [16] Optical recognition of handwritten digits dataset. <http://archive.ics.uci.edu/ml/machine-learning-databases/optdigits/>.
- [17] Seeds dataset. <http://archive.ics.uci.edu/ml/machine-learning-databases/00236/>.
- [18] Image segmentation dataset dataset. <http://archive.ics.uci.edu/ml/machine-learning-databases/image/>.

- [19] Breast cancer wisconsin dataset. <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>.

Поступила в редакцию 24.02.2016

## Metric learning and dimensionality reduction in clustering\*

*R. V. Isachenko and A. M. Katrutsa*

isa-ro@yandex.ru, amkatrutsa@yandex.ru

Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Moscow, Russia

This paper investigates incorporation of metric learning approach in clustering problem. Distance metric is a key issue in many machine learning algorithms, especially in unsupervised learning where distance between objects is the only known information. The metric learning procedure modifies distances between objects to make objects from the same cluster closer and from the different clusters more distant. In this paper, Mahalanobis distance is used as a distance between objects. The goal of the paper is to learn Mahalanobis metric by optimizing the covariance matrix of objects according to their cluster labels. In this case, metric learning procedure is formulated as optimization problem. For clustering,  $k$ -means were used as baseline algorithm and Adaptive Metric Learning (AML) algorithm. To solve the problem, AML algorithm uses iterative EM (expectation-maximization) procedure to find the optimum. To compare these algorithms, the computational experiment was carried out in MatLab on synthetic data and real data from UCI repository and conclusions about performance of these algorithms have been made.

**Keywords:** *clustering; k-means; adaptive distance metric learning; EM-algorithm*

**DOI:** 10.21469/22233792.2.1.02

## References

- [1] Yang, L., and R. Jin. 2006. Distance metric learning: A comprehensive survey. Michigan State University. Vol. 2. 51 p.
- [2] Bellet, A., A. Habrard, and M. Sebban. 2013. A survey on metric learning for feature vectors and structured data. arXiv:1306.6709.
- [3] Guillaumin, M., J. Verbeek, and C. Schmid. 2009. Is that you? Metric learning approaches for face identification. *IEEE 12th Conference (International) on Computer Vision Proceedings*. 498–505.
- [4] Yang, L., R. Jin, R. Sukthankar, and Y. Liu. 2006. An efficient algorithm for local distance metric learning. *AAAI* 2:543–548.
- [5] Globerson, A., and S. T. Roweis. 2005. Metric learning by collapsing classes. *Advances in neural information processing systems*. Eds. Y. Weiss, B. Schölkopf, and J. Platt. Cambridge, MA, USA: MIT Press. 18:451–458.
- [6] Xing, E. P., A. Y. Ng, M. I. Jordan, and S. Russell. 2003. Distance metric learning with application to clustering with side-information. *Advances in Neural Information Processing Systems* 15:505–512.
- [7] Geurts, P. 2001. Pattern extraction for time series classification. *Principles of data mining and knowledge discovery*. Springer. 115–127.

---

\*The work was supported by the Russian Foundation for Basic Research (grant No.16-37-00485).



- [8] Friedl, M. A., and C. E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61(3):399–409.
- [9] Hyafil, L., and R. L. Rivest. 1976. Constructing optimal binary decision trees is np-complete. *Inform. Proc. Lett.* 5(1):15–17.
- [10] Dy, J. G., and C. E. Brodley. 2004. Feature selection for unsupervised learning. *J. Machine Learning Res.* 5:845–889.
- [11] Ye, J., Z. Zhao, and H. Liu. 2007. Adaptive distance metric learning for clustering. *IEEE Conference on Computer Vision and Pattern Recognition.* 1–7.
- [12] MacQueen, J., *et al.* 1967. Some methods for classification and analysis of multivariate observations. *5th Berkeley Symposium on Mathematical Statistics and Probability Proceedings.* 1(14):281–297.
- [13] Ding, C., X. He, and H. D. Simon. 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. *SIAM Data Mining Conference Proceedings.* 606–610.
- [14] Shawe-Taylor, J., and N. Cristianini. 2004. *Kernel methods for pattern analysis.* Cambridge University Press. 478 p.
- [15] Letter recognition dataset. Available at: <http://archive.ics.uci.edu/ml/machine-learning-databases/letter-recognition/> (accessed March 28, 2016).
- [16] Optical recognition of handwritten digits dataset. Available at: <http://archive.ics.uci.edu/ml/machine-learning-databases/optdigits/> (accessed March 28, 2016).
- [17] Seeds dataset. Available at: <http://archive.ics.uci.edu/ml/machine-learning-databases/00236/> (accessed March 28, 2016).
- [18] Image segmentation dataset dataset. Available at: <http://archive.ics.uci.edu/ml/machine-learning-databases/image/> (accessed March 28, 2016).
- [19] Breast cancer wisconsin dataset. Available at: <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/> (accessed March 28, 2016).

*Received February 24, 2016*