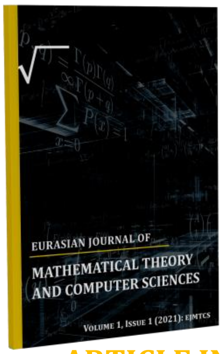
EURASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
Innovative Academy Research Support Center
UIF = 8.3 | SJIF = 5.916 www.in-academy.uz



ОБ ОДНОМ СПОСОБЕ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ

Рахманов А.Т.
Ташкентский университет информационных технологий
Ферганский филиал
Султонов С.
Ташкентский университет информационных технологий,
Ферганский филиал
https://www.doi.org/10.5281/zenodo.8385967

ARTICLE INFO

Received: 21thSeptember 2023
Accepted: 27thSeptember 2023
Online: 28th September 2023

KEY WORDS

Распознавание образов, классификация, способ, набор данных, обучающая выборка, алгоритм.

ABSTRACT

В работе исследована задача классификации для некоторых классов объектов с нелинейной поверхностью разделения. Решена задача классификации и предложен способ распознования, основанного на решении специальных систем неравенств, также построен соответствующий алгоритм классификации.

1. Введение

Классификация ⍰ это разделение объектов по различным двум классам на основе близости их параметров в пространстве признаков. Объектом понимается процесс или предмет (сущность) с заданным набором признаков. Решение задачи классификации это совокупность метода и соответствующего решающего правила и алгоритма для распределения объектов по классам.

Исследованию различных задач распознавания образов и классификации [1-4] посвящено много работ, например, [1-8]. При решении задачи классификации с нелинейной поверхностью разделения, возникает определенные трудности, связанные со структурой классифицирующих множеств. В этом случае для некоторых классов объектов не всегда возможно применение методов, предназначенных решении задачи классификации с линейной поверхностью разделения, например, [4-5,7-8]. В настоящей работе предлагается один способ решения задачи классификации, охватывающий различные классы объектов с нелинейной поверхностью разделения.

2. Постановка задачи. Рассмотрим обучающую выборку объектов $X=\{x\}$, каждый объект которого задается набором из n признаков $x=\{x^1,....,x^n\}$, $n>1$ и разбитую на непересекающиеся подмножества (классы) X_1, X_2. Пусть каждый класс X_p содержит m_p объектов $x_{p1},....,x_{pm_p}$, где $x_{pi}=(x^1_{pi},x^2_{pi},...,x^n_{pi})$, $i=\overline{1,m_p}$, $p=1,2$.

Далее, предподлжим, что определён [3,6]минимальный размер признакого пространства в смысле конкретно заданного критерия качества распознавания. Теперь постановку задачи классификации можно формулировать следующим образом: на построенном пространстве признаков осуществить классификацию объектов.

Для решения задачи осуществим дополнительные построения. Пространства признаков $x=(x^1,...,x^n)$ будем считать евклидовым и обозначим через R^n. Один из множеств X_1, X_2 и пусть это будет X_2, делим на k частей $X_{21}, X_{22}, \dots, X_{2k}$, так чтобы $X_2 = X_{21} \cup X_{22} \cup \ldots \cup X_{2k}$. Ясно такое деление всегда возможно, так как X_2 состоит из конечного числа элементов.

Среднеквадратичный разброс объектов в классе X_{2r} в R^n определяется формулой

$$d_r(x_{2r}) = \sqrt{\frac{1}{m_{2r}} \sum_{i=1}^{m_{2r}} \| x_{2r} - x_i \|^2},$$

где x_{2r} – некоторый элемент множества coX_{2r}, $x_i \in X_{2r}$, $i = 1,2,\dots,m_{2r}$, m_{2r} - количество элементов множества X_{2r}, $r = 1,2,\dots,k$. Ясно, что функция $d_r(x_{2r})$ является непрерывной на coX_{2r} по x_{2r} и поэтому на этом множестве существует минимальное значение данной функции. Пусть минимум функции $d_r(x_{2r})$ достигается в точке $\bar{x}_{2r} \in coX_{2r}$.

3. Решающее правило для классификации.

По известным векторам $\bar{x}_{21}$, $\bar{x}_{22}$, ... , $\bar{x}_{2k}$, определенных согласно п.2, из соответствующих классов X_{21}, X_{22},... , X_{2k}, используя построение из [7], определим следующие множества

$$S_{1,1} = \left\{ x : |x - \bar{x}_1| \le \frac{1}{\rho_{X_{11}}} |x - \bar{x}_{21}| \right\}, \quad S_{1,2} = \left\{ x : |x - \bar{x}_1| \le \frac{1}{\rho_{X_{12}}} |x - \bar{x}_{22}| \right\}, \dots, \quad S_{1,k} = \left\{ x : |x - \bar{x}_1| \le \frac{1}{\rho_{X_{1k}}} |x - \bar{x}_{2k}| \right\}$$

$$S_{2,1} = \left\{ x : |x - \bar{x}_{21}| \le \frac{1}{\rho_{X_{21}}} |x - \bar{x}_1| \right\}, \quad S_{2,2} = \left\{ x : |x - \bar{x}_{22}| \le \frac{1}{\rho_{X_{22}}} |x - \bar{x}_1| \right\}, \dots, \quad S_{2,k} = \left\{ x : |x - \bar{x}_{2k}| \le \frac{1}{\rho_{X_{2k}}} |x - \bar{x}_1| \right\}$$

где $\rho_{X_{ij}}, i, j = 1,2,\dots, k$ – константы и ясно, что $\bar{x}_1 \in coX_1$, $\bar{x}_{21} \in coX_{21}$, ... , $\bar{x}_k \in coX_{2k}$. Также как в [7,8] можно доказать, что каждое из множеств $S_{1,2}$, $S_{1,3}$, ..., $S_{1,k}$, $S_{2,1}$, $S_{2,3}$,..., $S_{2,k}$ является шаром при $\rho_{X_{ij}} > 1$. Отметим, что константы $\rho_{X_{ij}} > 1$ выбираются достаточно близкими к единице. Если для классов множеств X_1, X_{21}, ... , X_{2k} их соответствующие выпуклые оболочки coX_1, coX_{21},..., coX_{2k} строго разделимые, то из определения множеств $S_{1,2}$, $S_{1,3}$, ..., $S_{1,k}$, $S_{2,1}$, $S_{2,3}$,..., $S_{2,k}$ следует, что точка χ является близким к X_1 чем к X_2, если она принадлежить всем множествам $S_{1,2}$, $S_{1,3}$, ..., $S_{1,k}$, и наоборот, точка χ является близким к X_2 если она принадлежить хотя бы одному из множеств $S_{2,1}$, $S_{2,3}$,..., $S_{2,k}$.

Таким образом, если для произвольного объекта из $w \in R^n$ выполнено следующая система неравенств (1):

$$\begin{cases} |w - \bar{x}_1| \le \dfrac{1}{\rho_{x_{11}}} |w - \bar{x}_{21}|, \\ |w - \bar{x}_1| \le \dfrac{1}{\rho_{x_{12}}} |w - \bar{x}_{22}|, \\ \text{-----------------------} \\ |w - \bar{x}_1| \le \dfrac{1}{\rho_{x_{1k}}} |w - \bar{x}_k|. \end{cases},$$

(1)

то вектор w является объектом класса X_1. В противном случае рассмотрим неравенства (2), (3), (4):

$$\left\{ |w - \bar{x}_{21}| \le \frac{1}{\rho_{x_{21}}} |w - \bar{x}_1| \right\}, \qquad (2)$$

$$\left\{ |w - \bar{x}_{22}| \le \frac{1}{\rho_{x_{21}}} |w - \bar{x}_1| \right\}, \qquad (3)$$

………………,

$$\left\{ |w - \bar{x}_{2k}| \le \frac{1}{\rho_{x_{2k}}} |w - \bar{x}_1| \right\}. \qquad (4)$$

Если выпольнено хотя бы один из неравенств (2), (3),…, (4) то вектор w является объектом класса X_2. Если вектор w не удовлетворяет ни одному из условий (1)-(4), то он считается нераспознанным.

При получении разрешающего правило (1), (2), (3),…, (4) предполагалось, что выпуклые оболочки coX_1, coX_{21}, … , coX_{2k} соответствующих классов строго разделимые множества. На самом деле во многих прикладных задачах распознавания может оказаться, что множество coX_1 может пересекаться с множествами coX_{21}, … , coX_{2k} в нескольких точках. Это не означает, что данный метод не применим к решению задачи классификации в этом случае, так как условие пересекаемости множества coX_1 с множествами coX_{21}, … , coX_{2k} никак ни влияет на выполнение решающих неравенств (1), (2), (3),…,(4). В этом случае, множества coX_1, coX_{21}, … , coX_{2k} необходимо определить пренебрегая некоторыми отдаленными элементами соответствующих классов X_i, что влечет за собой осуществление классификации с некоторой погрешностью.

4. Алгоритм распознавания.

Шаг 1. Моделирование обучающей выборки объектов, которая разбита на подмножества (классы) X_1, X_2. Каждый объект задается набором из n признаков $x = \{x^1,...., x^n\}$, $n > 1$. Пусть каждый класс X_i содержит m_i объектов $x_{i1},..., x_{im_i}$, где $x_{ij} = (x_{ij}^1,..., x_{ij}^n)$, $i = 1,2$, $j = \overline{1, m_i}$.

Шаг 2. Определение множеств X_{21}, … , X_{2k}, $X_2 = X_{21} \cup X_{22} \cup ... \cup X_{2k}$ и объектов $\bar{x}_1 \in coX_1$, $\bar{x}_{21} \in coX_{21}$, … , $\bar{x}_{2k} \in coX_{2k}$ согласно пункту 2.

Шаг 3. Определение параметров $\rho_{X_{ij}}$, $i, j = 1,2,\dots ,k$ соответствующих классов X_1, X_{21}, … , X_{2k}.

Шаг 4. Распознавание объекта W, применяя решающие правила (1), (2), (3),…,(4).

References:

1. Горелик, А. Л. Общая постановка задачи распознавания объектов и явлений. // Кибернетикаю – 1980. №6. – С. 72-75.
2. Журавлев, Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации. // Проблемы кибернетики. – 1978. Вып. 33. С. 5-68.
3. Камилов М. М., Нишанов А. Х. Итерационный метод формирования информативных наборов признаков в задаче распознавания образов. // Проблемы информатики и энергетики. – 1992. №5, – С. 3-6.
4. Нишанов А. Х. Построение решающего правила в пространстве информативных признаков классифицируемых объектов. // Проблемы информатики и энергетики. – 1992. №5-6. С.8-11.
5. Нишанов, А. Х., Чернова О. Ю. Построение решающего правила в подпространстве информативных признаков. // Тезисы докладов. Ташкент, 8-9 сентября, 1994. С. 23-24.
6. Рахманов А.Т., Акбаралиев Б.Б., Эргашев А.К. Об одном методе сокращения размерности объёма выборки в интеллектуальном анализе данных.// Проблемы информатики и энергетики. 2011, №1,С.76-79, ISSN: 2010-7242.
7. Рахманов А.Т., Акбаралиев Б.Б., Рузибоев О.Б., У.А.Хасанов. Об одном модифицированном способе решения задачи классификации // Кимёвий технология, назорат ва бошқарув, 2014, №4(58), С.85-91.
8. Рахманов А.Т., Рузибаев О.Б. К решению задачи классификации для трёх классов. Science and world. 2015. №2(18). Vol.1. p. 21-25.
9. Sultonov, S. (2023). MASHINALI O'QITISH TUSHUNCHASI VA MASHINALI O'QITISH JARAYONINING UMUMIY QADAMLARI. *Engineering problems and innovations*.
10. Хайдаров, А., Султонов, С., & Билолов, И. (2022). ВЛИЯНИЕ ОТЖИГА НА РАЗМЕРЫ КРИСТАЛЛИЧЕСКИХ ЛАМЕЛЕЙ ПОЛИКАПРАМИДА. *Theoretical aspects in the formation of pedagogical sciences*, *1*(7), 319-321.
11. Rajabov, M., Rajabova, X., & Sultonov, S. (2023). MAKIAVELLIAN SHAXSINING O'ZIGA XOS XUSUSIYATLARI VA UNI ADABIYOTLARDAGI TA'RIFI. *Engineering problems and innovations*.
12. Muhammadodilovna, U. M., & Ashurovna, K. S. (2022). O'ZBEKISTON RESPUBLIKASIDA YOSHLARNI BILIM OLISH UCHUN YARATILAYOTGAN IMKONIYATLAR, ILM-FAN VA INNOVASYALARGA QIZIQTIRISHNING NOAN'ANAVIY USULLARI. *O'ZBEKISTONDA FANLARARO INNOVATSIYALAR VA ILMIY TADQIQOTLAR JURNALI*, *1*(10), 139-144.
13. Sultonov, S. . (2023). IMPORTANCE OF PYTHON PROGRAMMING LANGUAGE IN MACHINE LEARNING. *International Bulletin of Engineering and Technology*, *3*(9), 28–30. Retrieved from https://internationalbulletins.com/intjour/index.php/ibet/article/view/1020