

А. Н. Тырсин, Б. А. Шпак (Екатеринбург, УрФУ; Челябинск, ЮУрГУ). **Задача классификации многомерных наблюдений с помощью разделяющего эллипсоида.**

УДК 519.711.3+519.2/.6

Резюме: Рассмотрена задача бинарной классификации наблюдений многомерных наблюдений с помощью разделяющего эллипсоида. При этом группируются данные только из одного кластера. Предложенный алгоритм бинарной классификации апробирован применительно к задачам медицинской диагностики.

Ключевые слова: Классификация, метод, многомерное пространство признаков, наблюдения, эллипсоид, фактор, риск.

В настоящее время широко распространены методы бинарной классификации наблюдений по обучающим выборкам, которые разделяют экспериментальные данные на два кластера с помощью гиперплоскостей или поверхностей в многомерном пространстве признаков [1, 2]. При этом у каждого кластера предполагается наличие некоторого центра, вокруг которого группируются наблюдения.

Однако в этом случае затруднена интерпретация наблюдений, расположенных на значительном расстоянии от прямой, проходящей через центры тяжести (середины) кластеров.

В отличие от указанных методов классификации предлагаемый подход позволяет разделить два класса по ограниченной геометрической поверхности — эллипсоиду. Процедура построения разделяющего эллипсоида можно реализовать различными способами. Например, путем минимизации расстояния между наблюдениями из двух кластеров до разделяющего эллипсоида.

В качестве примера рассмотрим классификацию пациентов на здоровых и нездоровых по показателям факторов риска $\mathbf{X} = (X_1, \dots, X_m)$.

Кластер здоровых пациентов группируется относительно некоторого центра $\mu = (\mu_1, \dots, \mu_m)$. А нездоровые пациенты могут иметь самые разнообразные отклонения значений диагностических признаков и плохо группироваться. Однако точки, характеризующие значения факторов риска не здоровых пациентов, удалены от этого центра. Поэтому, перенеся этот центр в начало координат, можно осуществить классификацию пациентов на здоровых и не здоровых по обучающей выборке с помощью разделяющего эллипсоида

$$G = \left\{ \mathbf{x} = (x_1, \dots, x_m) : \sum_{j=1}^m \frac{(x_j - \mu_j)^2}{B_j^2} = 1 \right\}.$$

В этом случае расположение конкретных значений признаков не имеет значения, так как критерием эффективности будет расстояние точки до поверхности эллипсоида или вероятность отнесения точки к внутренней части эллипсоида. Одновременно данный метод позволит формализовать процедуру настройки пороговых уровней и задания функции риска в модели многомерного риска [3]. А именно, по каждому фактору риска X_k пороговые уровни будут равны $\mu_k \mp B_k$.

Данный метод классификации в задачах медицинской диагностики позволит повысить достоверность распознавания и интерпретации данных. Предложенный алгоритм бинарной классификации был апробирован применительно к задаче оценивания популяционного риска возникновения хронических не инфекционных заболеваний.

Исследование выполнено при финансовой поддержке гранта РФФИ, проект № 20-51-00001.

СПИСОК ЛИТЕРАТУРЫ

1. *Hosmer D. W., Lemeshow S., Sturdivant R. X.* Applied Logistic Regression. 3rd ed., Wiley 2013.
2. *McLachlan G. J.* Discriminant Analysis and Statistical Pattern Recognition, Wiley, 2004.
3. *Тырсин А. Н., Сурина А. А.* Моделирование риска в многомерных стохастических системах. — Вестник Томского государственного ун-та. Управление, вычислительная техника и информатика. 2017, № 2(39), с. 65–72.

УДК 519.711.3+519.2/.6

Tyrsin A. N., Shpakov B. A. (Yekaterinburg, Russia, Ural Federal University; Chelyabinsk, Russia, South Ural State University). **The problem of classifying multidimensional observations using a dividing ellipsoid.**

Abstract: The problem of binary classification of observations of multidimensional observations using a separating ellipsoid is considered. However, we only group data from one cluster. The proposed binary classification algorithm has been tested in relation to the problems of medical diagnostics.

Keywords: Classification, method, multidimensional parameter space, observations, dividing ellipsoid, factor, risk.