

Гасанов Александр Русланович,
Тюменский государственный университет,
Институт математики и компьютерных наук,
студент МОиАИС-174,
alexgasanov89@gmail.com

Сизова Людмила Владимировна,
Тюменский государственный университет,
Институт социально-гуманитарных наук,
Кафедра иностранных языков и межкультурной
профессиональной коммуникации,
старший преподаватель,
l.v.sizova@utmn.ru

МЕТОД ОКНА ПАРЗЕНА

Alexander R. Gasanov,
the University of Tyumen,
Institute of Mathematics and Computer Science,
student of MSaISA-174,
alexgasanov89@gmail.com

Lyudmila V. Sizova,
the University of Tyumen,
Department of Foreign Languages and
Intercultural Professional Communication ,
senior lecturer,
l.v.sizova@utmn.ru

PARZEN WINDOW

Аннотация: В статье рассматривается один из методов решения задачи классификации. В век развития искусственного интеллекта все больше задач становятся разрешимы методами машинного обучения. С помощью информатизации в различных сферах деятельности человека, появляются огромные массивы данных, с помощью которых можно прогнозировать различные события. Методы машинного обучения хорошо с этим справляются и помогают искать связи в данных. В статье представлена информация о решении задачи классификации методом окна Парзена. Авторы статьи надеются, что читатели получат больше информации по данной проблеме и узнают, как применяются методы машинного обучения в нашем мире.

Ключевые слова: машинное обучение, задача классификации, метод ближайших соседей, метод окна Парзена.

Abstract. The article considers one of the methods of solving classification problem. In the century of developing artificial intelligence more and more problems become solvable by machine learning methods. With the help of informatization in different spheres of human's life there are a lot of big data arrays, which can help to predict different events. Machine learning methods handle it well and help to find out relations in data. The article provides the information about solving classification problem with Parzen window method. The authors of the article hope the readers can be better informed of the problem and how machine learning methods can be used in our world.

Key words: machine learning, classification problem, nearest neighbors, Parzen window method.

Barack Obama won the election thanks to artificial intelligence. He had a team of data analysts, who predicted an actual topic for the presentation in every state. They did this with the help of machine learning. Today you can predict anything using machine learning methods.

Machine learning engineers often meet classification problem. It is necessary to understand, what it is. Imagine, you are an analyst and you should predict who will win the election. What would you do? The first good idea is to take a poll of the citizens. Do we really need to ask every citizen? It will take a lot of time. Now imagine that we have some information about citizens such as address, number of children, age, salary etc. We will call it arguments. Some of citizens would say who will get their vote on election. We will call it a marker. Using this information, we need to find a rule, which divides people into political groups. This rule depends on arguments. We have got a function which marks people using information about their life. That is an example of the classification problem. We have got objects with arguments and need to mark them knowing information about arguments and real markers of some group of objects. We will call this group a training set. It is not a definite method, it is just approximation. But this method could give good predictions, where the percent of mistakes is very little.

The next question: How to classify objects? ML engineers often use metrical algorithms. The main idea of metrical classifier is analysis of distances between objects. The classical example is KNN algorithm. We are going back to election. That sounds logical that people in ghetto will elect the candidate who wants everyone to be financially equal. Businessmen will elect the candidate who wants to promote capitalism. If we ignore other arguments and compare citizens' salary, we determine that people with the same salary will elect the same candidate. Now we add the second argument – number of children. First candidate wants to lower taxes for people with children and higher taxes for single people. We have got a coordinate plane. We assume that a random citizen

will have the same mark as most of his 5 nearest neighbors. That is the easiest metrical classifier. We should count distances between objects from a training set and objects from a test set. Then determine that the most of N nearest objects have the same mark.

Another example of metrical algorithm is Parzen window classifier. It is a metrical algorithm, where nearest neighbors have got a weight. We are backing to citizens. Now we have got 10 arguments for each citizen. It is logical, that a man with 7 common parameters is more similar than a man with 4 common parameters. Imagine, that you have got 3 the most similar people in the city. The first person has 7 common parameters with you, the second one has 5 parameters, and the third has 3 common parameters. All of them have a weight. The nearest neighbor has the biggest weight. The farthest has the least weight. We should calculate total weight of every candidate. The candidate with the biggest total weight is most likely to take our vote on. We have got nonincreasing function of weights, which gives an importance for each of the nearest neighbors.

References:

1. Regression and classification. (2018). Available from: <https://www.geeksforgeeks.org/regression-classification-supervised-machine-learning/> [Accessed 20/02/2019].
2. Metod Parzenovskogo okna [Method of Parzen window]. Available from: https://studme.org/139985/informatika/metod_parzenovskogo_okna/ [Accessed 22/03/2019].