**Перцептрон в задачах бинарной классификации**

**Митина Ольга Алексеевна**

кандидат пед. наук.,

МИРЭА – Российский технологический университет,

г. Москва

**Ломовцев Павел Павлович**

студент 2 курса,

МИРЭА – Российский технологический университет,

г. Москва

**PERCEPTRON IN BINARY CLASSIFICATION PROBLEMS**

**Mitina Olga Alekseevna**

Candidate of Science

MIREA – Russian Technological University

Moscow

**Lomovtsev Pavel Pavlovich**

2nd year student

MIREA – Russian Technological University

Moscow

**Аннотация**

В настоящее время в России растет объем продукции, производимой предприятиями. Одной из серьезных проблем для российских предприятий является создание системы автоматического определения категории товара, позволяющей обеспечить безошибочную сортировку различных объектов. Очередной товар (объект) необходимо отнести к верному классу в зависимости от его характеристик.

Перцептрон – один из наиболее популярных методов решения задачи классификации. Автоматизация процесса разделения товаров по их свойствам облегчит работу сотрудникам сортировочного центра и исключит возможность ошибки вследствие человеческого фактора.

**Ключевые слова**: перцептрон; задачи бинарной классификации; нейронные сети.

**ANNOTATION**

Currently, the volume of products manufactured by enterprises is growing in Russia. One of the serious problems for Russian enterprises is the creation of a system for automatic determination of the category of goods, which makes it possible to ensure error-free sorting of various objects. The next product (object) must be attributed to the correct class, depending on its characteristics. The perceptron is one of the most popular methods for solving the classification problem. Automation of the process of separating goods by their properties will facilitate the work of the employees of the sorting center and eliminate the possibility of errors due to human factors.

**Keywords:** perceptron; binary classification tasks; neural networks.

Классификация – один из разделов машинного обучения, посвященный решению задачи на множестве объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Классовая принадлежность остальных объектов не известна.

**Бинарная классификация** – разбиение исходного множества объекта на **два** различных подмножества таким образом, чтобы характеристики объектов отличались друг от друга, но внутри каждого из подмножеств свойства были схожи [1].

Задача бинарной классификации – одна из задач **обучения с учителем** (англ. Supervised learning). Это означает, что для того, чтобы алгоритм работал правильно, ему нужны объективные данные. На каждом шаге результат работы алгоритма будет сравниваться с уже существующим единственно верным результатом и если он окажется неверным, необходимо корректировать параметры классификатора до тех пор, пока не получится верный результат.

Задача бинарной классификация требует внимательного подхода к выбору алгоритма, а также к его обучению (корректировки параметров).

Мы будем использовать один из алгоритмов бинарной классификации – перцептрон, который моделирует работу человеческого мозга.

По определению перцептрон – математическая модель восприятия информации головным мозгом, предложенная Фрэнком Розенблаттом в 1957 году. Перцептрон стал одной из первых нейросетей в мире [2].

Перцептрон состоит из трёх типов элементов, а именно: поступающие от [датчиков](https://ru.wikipedia.org/wiki/%D0%94%D0%B0%D1%82%D1%87%D0%B8%D0%BA) [сигналы](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%B3%D0%BD%D0%B0%D0%BB) передаются [ассоциативным](https://ru.wikipedia.org/wiki/%D0%90%D1%81%D1%81%D0%BE%D1%86%D0%B8%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%B0%D1%8F_%D0%BF%D0%B0%D0%BC%D1%8F%D1%82%D1%8C) элементам, а затем реагирующим элементам [2].

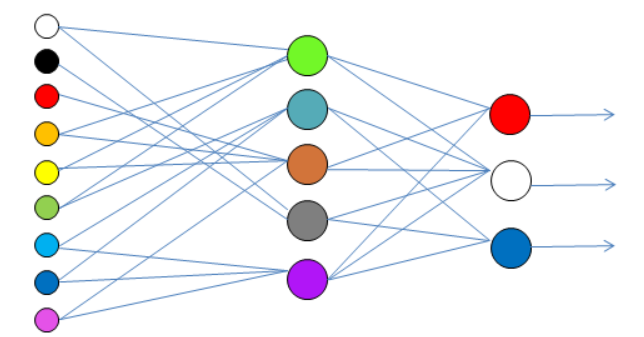


Рисунок 1 Логическое представление перцептрона с тремя уровнями

На рисунке 1 показана логическая схема перцептрона с тремя выходами, которая показывает, каким образом входные данные преобразуются в выходные, проходя обучение через один А-слой.

Алгоритм широко применяется в различных сферах: в медицине, в информационной безопасности, классификации товаров по качеству, в криминальном деле, в задачах распознавания изображений.

Перцептроны позволяют создать набор [«ассоциаций»](https://ru.wikipedia.org/wiki/%D0%90%D1%81%D1%81%D0%BE%D1%86%D0%B8%D0%B0%D1%86%D0%B8%D1%8F_(%D0%BF%D1%81%D0%B8%D1%85%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F)) между входными [стимулами](https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B8%D0%BC%D1%83%D0%BB) и необходимой [реакцией](https://ru.wikipedia.org/wiki/%D0%A0%D0%B5%D0%B0%D0%BA%D1%86%D0%B8%D1%8F_(%D0%B1%D0%B8%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F)) на выходе. В биологическом плане это соответствует преобразованию, например, зрительной информации в [физиологический ответ](https://ru.wikipedia.org/w/index.php?title=%D0%A4%D0%B8%D0%B7%D0%B8%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D1%80%D0%B5%D0%B0%D0%BA%D1%86%D0%B8%D1%8F&action=edit&redlink=1) от двигательных [нейронов](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D1%8B).

Для оценки качества классификации необходимо рассчитать все необходимые метрики, которые позволят определить, насколько хорошо модель предсказывает класс очередного объекта.

Под метрикой в задачах машинного обучения понимают количественный показатель (в долях), который показывает ту или иную характеристику классификатора [4].

Метод оценки качества классификации – инструментарий для количественной оценки результатов классификации [3].

Принято выделять две группы методов оценки качества классификации:

* внешние (англ. External) меры основаны на сравнении результата классификации с априори известным разделением на классы;
* внутренние (англ. Internal) меры отображают качество кластеризации только по информации в данных [5].

Мы будем использовать внешнюю меру оценки качества классификации, т. е. по данным с априори известным реальным распределением данных.

В настоящее время специалисты по машинному обучению используют следующие метрики для оценки качества классификации: accuracy, precision, recall, F-мера [6].

Принято считать, что модель должна быть сбалансированной, то есть данные метрики должны быть не менее 0.7 на тестовом наборе данных – наборе, который модель видит впервые, иначе модель будет плохо классифицировать новые объекты.

Итак, классификация по двум классам, или бинарная классификация, является едва ли не самой распространенной задачей машинного обучения. Рассмотрим пример бинарной классификации на наборе данных, в которых присутствует бинарный признак.

В нашем случае такой столбец описывает наименование цитруса:

* «1» – апельсин;
* «-1» – грейпфрут.

На рисунке 2 представлены данные, на которых будет производится тестирование, где:

* name – целевой столбец факторного типа данных;
* diameter – диаметр фрукта, числового типа данных;
* weight – вес фрукта, числового типа данных;
* red – оттенок красного у фрукта, числового типа данных;
* green – оттенок зеленого у фрукта, числового типа данных;
* blue – оттенок синего у фрукта, числового типа данных. [8]

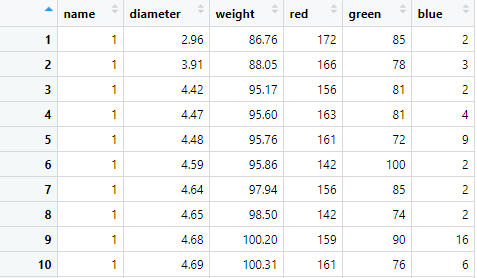


Рисунок 2 Описание тренировочных данных

Для реализации перцептрона будем использовать язык программирования высокого уровня R для статистической обработки данных и работы с графикой, а также свободную программную среду вычислений с открытым исходным кодом в рамках проекта GNU [6].

В настоящее время язык R используется в крупных компаниях для анализа, визуализации данных, а также для сложных математических операций. Преимуществом этого языка программирования является то, что все операции векторизованы, что исключает явное использование циклов в программном   
коде [7].

Импортировав данные в рабочую среду R, разделим исходные данные на тренировочную и тестовую выборку. Тренировочная – для обучения модели, тестовая – для тестирования.

Таким образом, классификация данных – задача определения, к какому классу из известных относится новый объект. Существует большое количество примеров применения алгоритмов классификации, такие как: задачи медицинской диагностики, предсказание месторождений полезных ископаемых, оценка кредитоспособности заемщиков, то есть случаи, где решения принимаются на основании принадлежности данных к одному из выделенных классов. Благодаря этим технологиям жизнь может стать проще, а принимаемые решения – точнее и дешевле.

**Список литературы**

1. Горяинова, Е.И. Методы бинарной классификации объектов с номинальными показателями / Е.И. Горяинова // Журнал Новой экономической ассоциации № 2 (14), C. 27–49. – 2017. – С. 35
2. Warren S. McCulloch and Walter Pitts. [Логическое исчисление идей, относящихся к нервной активности](http://www.raai.org/library/books/mcculloch/mcculloch.pdf) = [A logical calculus of the ideas immanent in nervous activity](http://www.springerlink.com/content/61446605110620kg/fulltext.pdf) // Bulletin of Mathematical Biology. – New York: Springer New York, 1943. – Т. 5, № 4. – С. 115—133.
3. Жуков, Д.А. Анализ критериев классификации при диагностике функционирования технического объекта / Д.А. Жуков // Mathematical modelling. – 2018. – С. 13
4. Роберт Кабаков. R в действии. – [ДМК-Пресс](https://ru.wikipedia.org/w/index.php?title=%D0%94%D0%9C%D0%9A-%D0%9F%D1%80%D0%B5%D1%81%D1%81&action=edit&redlink=1), 2018. – 588 с.
5. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности. – Москва: Финансы и статистика, 2019 – 571 с
6. Habr.com [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/ods/blog/328372/> – Дата доступа: 17.04.2021.
7. Мастицкий, С.Э. Статистическй анализ и визуализация данных с помощью R / С.Э. Мастицкий. – Москва: 2017. – 172 с.
8. Kaggle.com [Электронный ресурс]. – Режим доступа: https://www.kaggle.com/joshmcadams/oranges-vs-grapefruit?select=citrus.csv. – Дата доступа: 17.04.2021.