

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

**«Дальневосточный федеральный университет»**

**(ДВФУ)**

|  |  |
| --- | --- |
|  | **ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ** |

**Лабораторная работа №10**

**Наивный Байесовский классификатор**

Дисциплина «Теория вероятностей и математическая статистика»

Студент группы Б9123-01.03.02ии

Моттуева Уруйдана Михайловна

г. Владивосток

2025

Целью данной лабораторной работы является построение модели наивного Байесовского классификатора в виде реализации класса NaiveBias.

**Постановка задачи классификации**

Рассматривается генеральная совокупность объектов с признаками

. Каждому объекту присвоен один из нескольких классов .

Требуется на основе выборки определить, к какому классу относится объект, отсутствующий в выборке.

**Наивный Байесовский классификатор**

**Идея**

Если новый объект имеет значения признаков , то можно вычислить условную вероятность принадлежности к классу :

Где:

1. – это условная вероятность того, что событие произойдет при условии, что событие уже произошло.
2. – это априорная вероятность события Показывает, какова вероятность события ​ без учета информации о *X*.
3. – это условная вероятность события при условии, что событие ​ произошло. Показывает, насколько вероятно наблюдение , если известно, что ​.
4. – вероятность события . Представляет собой общую вероятность события , независимо от значения .

**Проблема**

Неизвестны ни частные, ни совместные, ни условные функции вероятности.

**Оценка функций вероятности**

Для функции вероятности можно использовать несмещённую состоятельную оценку — относительная частота значения у признака .

Аналогично для оценка — относительная частота класса y.

Для условной функции вероятности оценка — относительная частота значения у признака среди объектов класса .

**Наивное предположение**

Для вычисления совместной функции вероятности предположим, что признаки независимы в совокупности. Тогда

**Принятие решения о классе**

Одно из возможных правил выбора класса — выбрать класс с наибольшей вероятностью.

Тогда для объекта со значениями признаков будет выбран класс

Можно заметить, что не зависит от , а вместо вероятностей можно брать их логарифм. Следовательно, итоговый вид модели:

**Замечания о модели**

1. Для непрерывных признаков следует использовать плотность нормального распределения:

где:

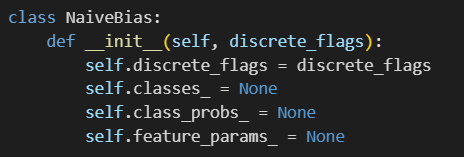
- — среднее арифметическое,

- — выборочное среднеквадратическое отклонение значений признака у объектов класса .

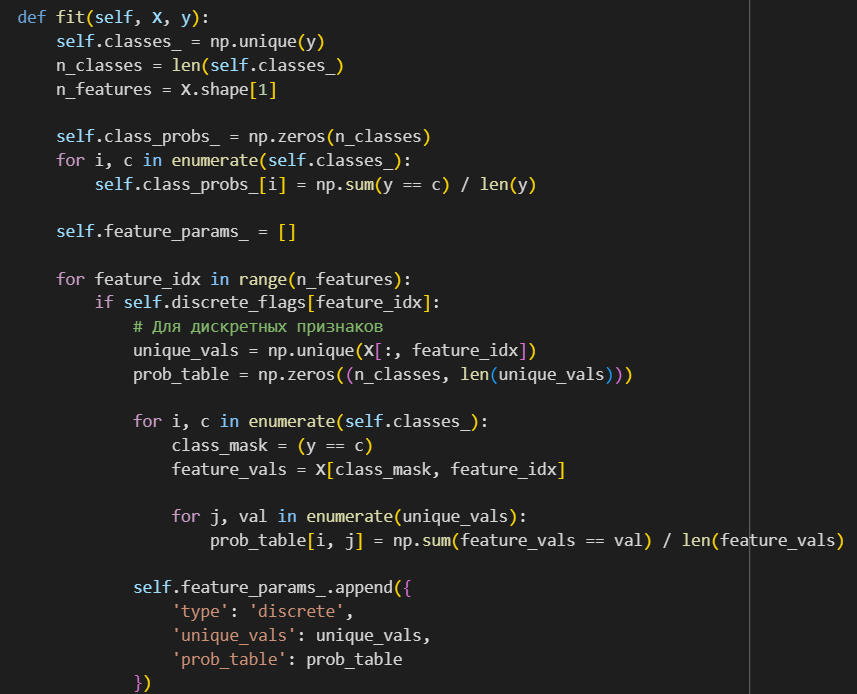
2. Чтобы относительная частота значения не равнялась 0, следует при её вычислении прибавить 1 к числителю и знаменателю.

Реализация

1. Конструктор принимает флаги, отвечающие за то, какие признаки являются дискретными, а какие непрерывными;



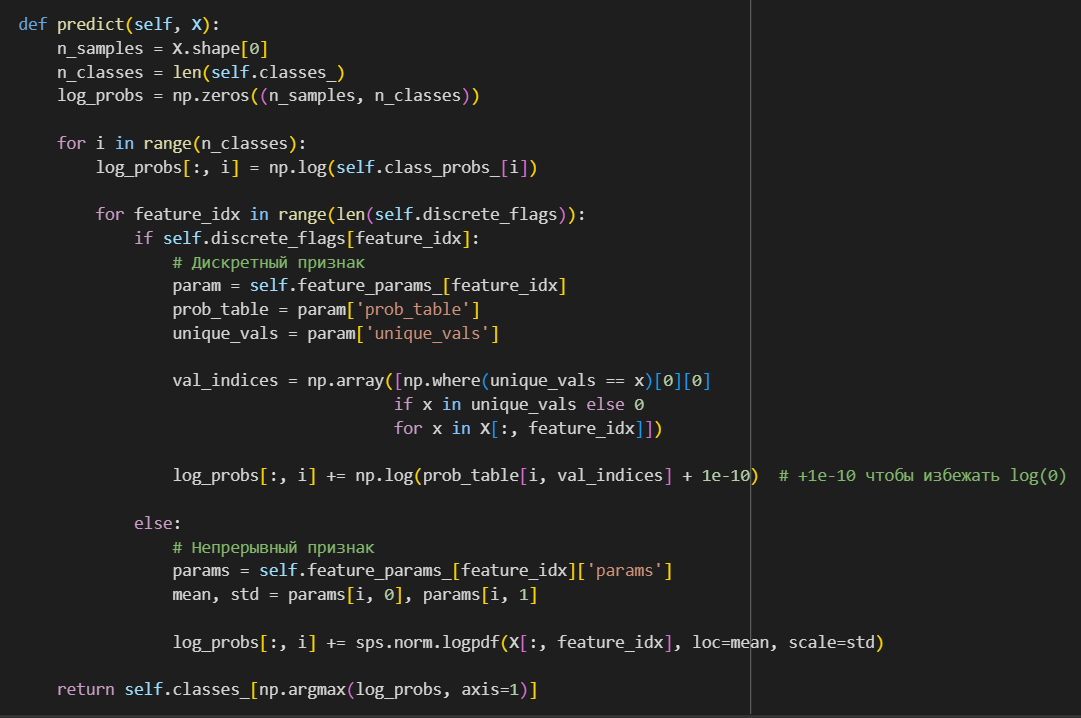
1. Метод fit принимает массив объектов X и вектор классов y из тренировочной выборки. Вычисляет и сохраняет необходимые значения функций вероятности и плотностей;



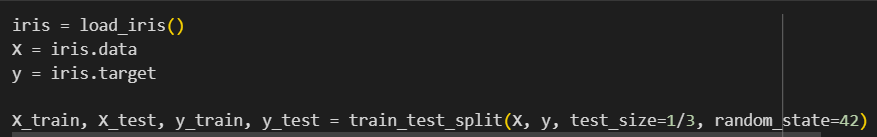
Изображение выглядит как текст, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.

1. Метод predict принимает массив объектов X и возвращает вектор предсказанных классов.

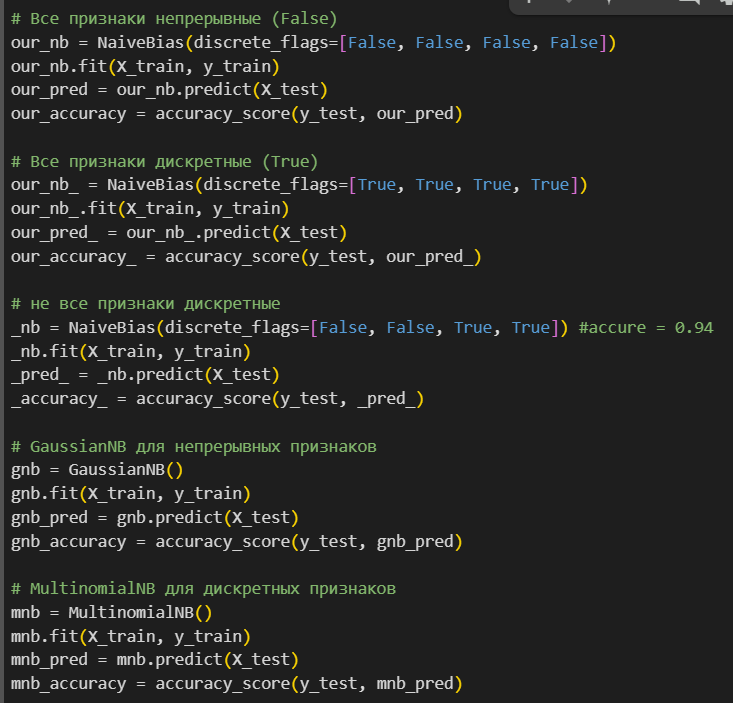


Разбейте выборку на тренировочную и тестовую, указав долю тестовой выборки, например 1/3, и выбрав случайное число random\_state для перемешки выборки.



Изображение выглядит как текст, снимок экрана, Шрифт

Контент, сгенерированный ИИ, может содержать ошибки.



Изображение выглядит как текст, Шрифт, снимок экрана

Контент, сгенерированный ИИ, может содержать ошибки.