

Национальный исследовательский университет
компьютерных технологий, механики и оптики

Факультет ПИиКТ

Системы искусственного интеллекта.
Лабораторная работа №3.
Вариант 11 (нечетный).

Работу выполнил: Кулаков Н. В.
Группа: Р33312
Преподаватель: Королева Ю. А.

Санкт-Петербург
2022

1 Описание задания

1. Датасет с данными про оценки студентов инженерного и педагогического факультетов (для данного датасета нужно ввести метрику: студент успешный/неуспешный на основании грейда)
2. Отобразить случайным образом \sqrt{n} признаков.
3. Реализовать без использования сторонних библиотек построение дерева решений (numpy и pandas использовать можно).
4. Провести оценку реализованного алгоритма с использованием Accuracy, precision и recall
5. Построить AUC-ROC и AUC-PR.

2 Реализация.

2.1 Подсчет количества признаков

Общее количество признаков в датасете - 32. Тогда требуемое количество:

$$k = \lceil \sqrt{32} \rceil = 6$$

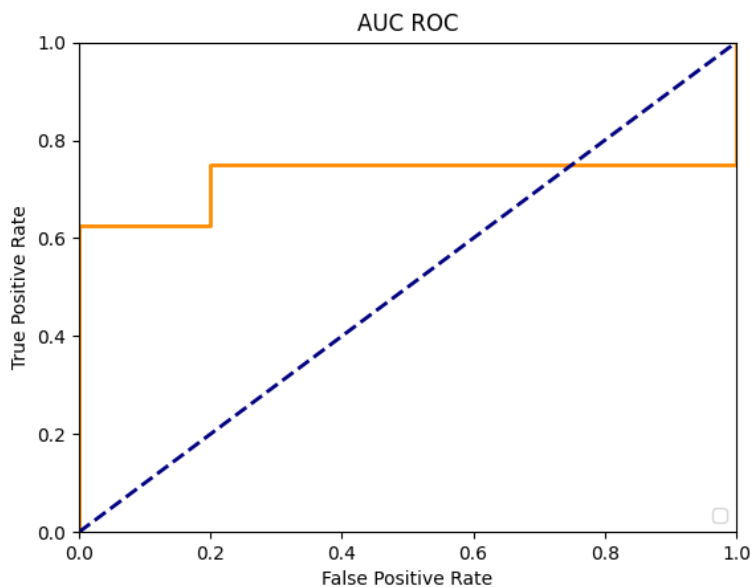
2.2 Реализация и вывод формул

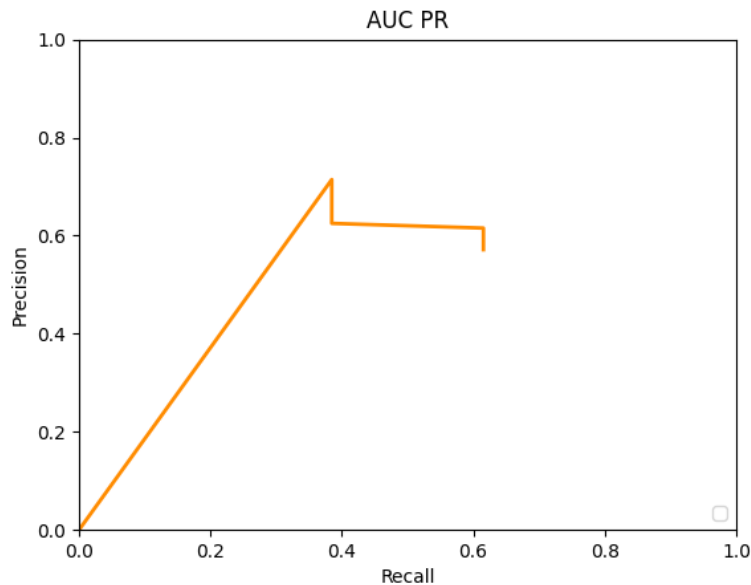
Обратиться к пункту 3.

2.3 Оценка реализованного алгоритма с использованием accuracy, precision и recall

Accuracy: 0.6551724137931034
Precision: 0.6153846153846154
Recall: 0.6153846153846154

2.4 AUC-ROC и AUC-PR





3 Исходный код

Результаты, полученные в ходе выполнения лабораторной работы: [sources](#).

4 Вывод.

В ходе выполнения лабораторной работы я познакомился с алгоритмом C4.5 и реализовал его на питоне. Также вспомнил, что такое accuracy, precision, recall, а также разобрался, как строить графики AUC-ROC, AUC-PR.

Графики AUC-PR, AUC-ROC довольно неинформативны, поскольку размер входных данных составляет 150 строк, тестовое разбиение 3:7. Именно поэтому количество данных, по которым, учится алгоритм, а также количество тестовых данных, мало, что свидетельствует о небольшом кол-ве точек на графике.