МИНОБРНАУКИ РОССИИ

Федеральное государственное автономное образовательное

учреждение высшего образования

«Южный федеральный университет»

Институт высоких технологий и пьезотехники

|  |  |
| --- | --- |
| Описание: https://pp.vk.me/c623122/v623122412/1acb7/yTJRs3eyFiE.jpg | Кафедра прикладной информатики и инноватики  Направление подготовки: 09.03.03 Прикладная информатика |

**Отчет на тему «Лабораторная работа №2: Деревья решений»**

**По дисциплине «Основы машинного обучения»**

**Выполнил:**

студент 3 курса бакалавриата 6 группы

Кузин М.А.

**Проверил:**

доцент, к. ф-м. н.

Ефимов А.И.

Ростов-на-Дону

2022

**Цели и задачи работы**

Изучить принцип работы метода машинной инженерии деревья решений для задач классификации. Реализовать на практике данный метод с помощью стандартных методов, библиотеки scikit-learn для Toy’s datasets классификации, создать с помощью pandas и numpy свою реализацию деревьев решений для заданных датасетов. Вывести графики бинарных графов деревьев решений

**Использованные инструменты**

Python 3.9

Библиотеки: Sklearn(datasets, KNeighborsClassifier и т.д.), numpy, matplotlib(pyplot)

**Краткая теория**

Деревья решений (DT) — это непараметрический контролируемый метод обучения, используемый для классификации и регрессии . Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной, изучая простые правила принятия решений, выведенные из характеристик данных. Дерево можно рассматривать как кусочно-постоянное приближение.

Например, в приведенном ниже примере деревья решений обучаются на основе данных, чтобы аппроксимировать синусоидальную кривую с набором правил принятия решений «если-то-еще». Чем глубже дерево, тем сложнее правила принятия решений и тем лучше модель.

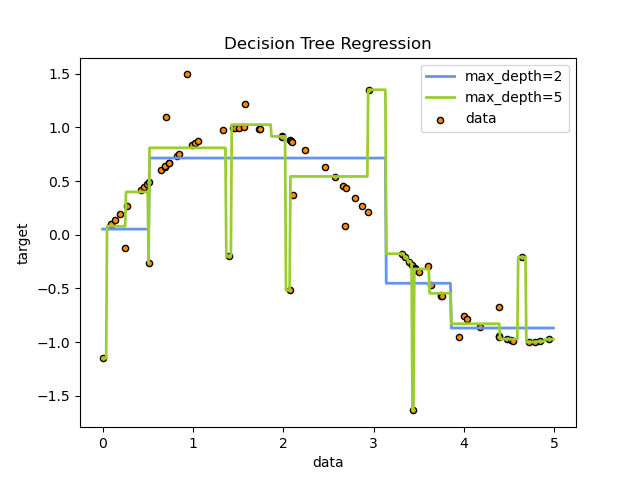


Рисунок – 1. Аппроксимация синусоидальной кривой с набором правил деревьев решений.

**Ход работы**

*Часть 1. Реализация деревьев решений с помощью Numpy и Pandas*

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 2. Методы энтропии и gini.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 3. Методы построения дерева решений (с помощью рекурсии).

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 3. Методы предсказания.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 4. Построение графика.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 5. Применение кросс-валидации и вывод рез-тов.

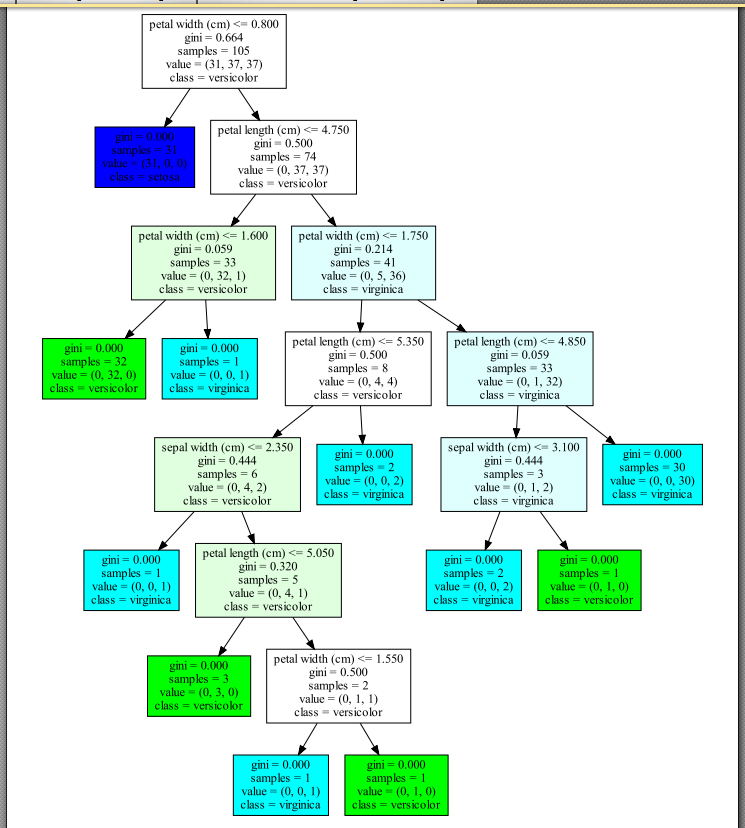


Рисунок - 6. Бинарный графа для Iris.

Изображение выглядит как текст, снимок экрана

Автоматически созданное описание

Рисунок - 7. Бинарный графа графа для Digits.

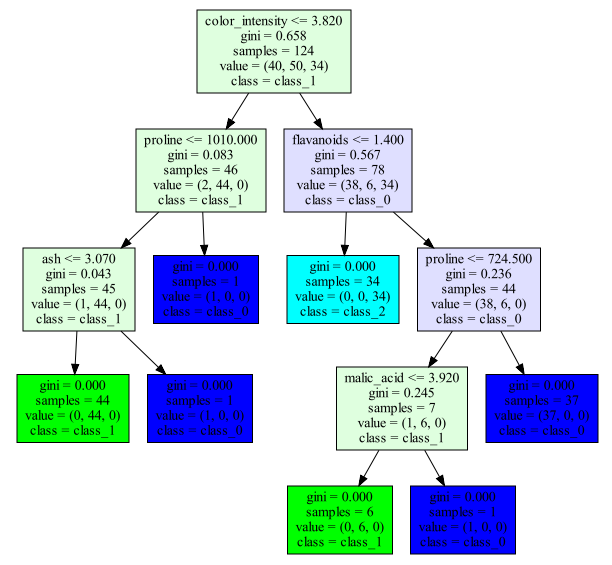


Рисунок - 8. Бинарный графа для Wine.

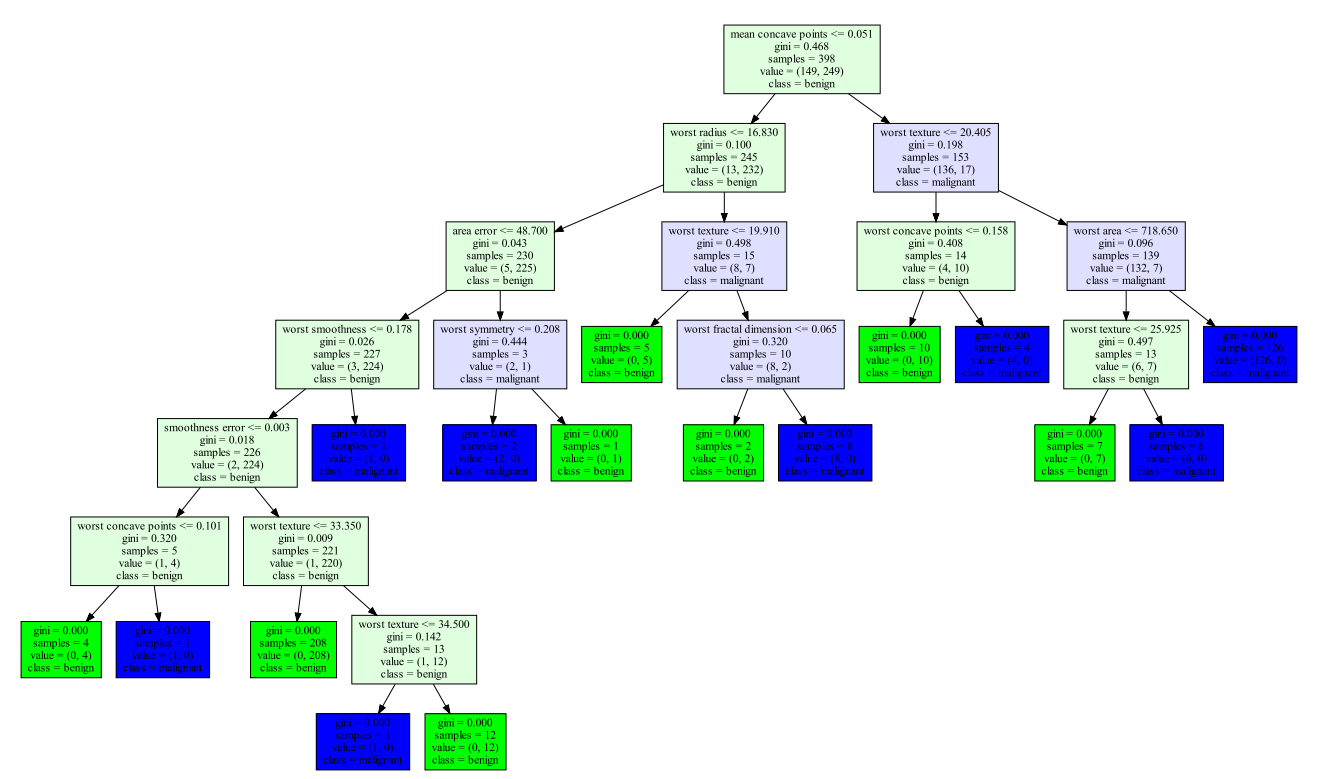


Рисунок - 9. Бинарный графа для Breast cancer wisconsin.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок - 10. Результаты обучения для Toy’s datasets.

**Вывод**

Изучили принцип работы метода машинной инженерии деревьев решений для задач классификации. Реализовали на практике данный метод с помощью стандартных функций, библиотеки scikit-learn для Toy’s datasets классификации. Также для этих датасетов, написали программу, которая строит бинарный граф на обученной модели дерево решений.