

Індивідуальне завдання: дерева рішень.

Встановіть значення змінної `variant`: сума номера групи помноженого на 25 і порядкового номеру студента в списку групи (групі ПМОм-11 відповідає номер 0, групі ПМІм-11 відповідає номер 1, групі ПМІм-12 відповідає номер 2, групі ПМІм-13 відповідає номер 3). Далі встановіть `set.seed(variant)` та згенеруйте значення змінної `redundant` як заокруглене до цілого (для заокруглення можна використати функції `floor` або `round`) випадкове число з рівномірного на інтервалі (номер групи + 5, 25 – номер групи) розподілу (функція `runif`).

1. Модифікуйте дані `Auto` наступним чином: встановивши `seed`, що дорівнює значенню змінної `variant`, видаліть `redundant %` спостережень з допомогою функції `sample`. Передбачте `mpg` на модифікованих даних `Auto` використовуючи дерева регресії. Поділіть дані на навчальний та тестовий набори. До тестового набору включіть `redundant %` спостережень. Пристосуйте дерево регресії до навчальних даних. Зобразіть дерево графічно та проінтерпретуйте отримані результати. Яку тестову помилку отримано? Використайте перехресну перевірку для визначення оптимального рівня складності дерева. Чи покращує обрізання тестову помилку? Використайте бутстрап агрегацію для аналізу даних. Яку тестову помилку отримано? Визначте, які змінні є найбільш важливі. Використайте випадкові ліси для аналізу даних. Яку тестову помилку отримано? Визначте, які змінні є найбільш важливі. Опишіть ефект вибору m , кількості змінних, що розглядаються при кожному поділі, на отриману тестову помилку. Застосуйте підсилення для навчальних даних для діапазону значень параметра стискання λ . Побудуйте графік з різними значеннями параметра стискання на осі x і відповідний навчальний MSE на осі y . Побудуйте графік з різними значеннями параметра стискання на осі x і відповідний тестовий MSE на осі y .