Індивідуальне завдання: класифікація.

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). Також згенеруйте значення змінної year як заокруглене до цілого випадкове число з рівномірного на інтервалі (2006, 2008) розподілу.

1. Використовуючи дані Weekly, встановивши seed, що дорівнює значенню змінної variant, побудуйте модель логістичної регресії для передбачення Direction з використанням навчальних даних з 1990 по year рр., на основі єдиного предиктора Lag2. Обчисліть матрицю помилок та загальну частку правильних прогнозів на тестових даних (тобто даних за year-2010 роки). За аналогічних умов використайте для передбачення Direction лінійний дискримінантний аналіз, квадратичний дискримінантний аналіз та метод K-найближчих сусідів з K = 1. Порівняйте використані методи. За якого K точність методу K-найближчих сусідів буде найбільшою?

2. Модифікуйте дані Auto наступним чином: встановивши seed, що дорівнює значенню змінної variant, видаліть redundant % спостережень з допомогою функції sample. Створіть двійкову змінну mpg01, яка містить 1, якщо mpg містить значення більше за середнє, і 0, якщо mpg містить значення менше за середнє. Розбийте дані на навчальний та тестовий набори. При розбитті набору даних **обов'язково**!!! встановити seed, що дорівнює значенню змінної variant, та використати функцію sample. Обсяг тестової вибірки виберіть 2*redundant % від загального обсягу даних. Застосуйте лінійний дискримінантний аналіз, квадратичний дискримінантний аналіз, логістичну регресію та метод K-найближчих сусідів з різними значеннями для K на навчальних даних, щоб передбачити mpg01 на основі змінних weight, displacement та horsepower. Порівняйте тестові помилка використаних моделей. За якого K точність методу K-найближчих сусідів буде найбільшою та якою буде точність цього методу?