1. Розглянемо дані Weekly, які є частиною пакету ISLR. Ці дані за своїм характером схожі на дані Smarket, за винятком того, що вони містять 1 089 щотижневі дохідності за 21 рік, з початку 1990 р. до кінця 2010 року.

1.1. На основі певних чисельних та графічних результатів щодо даних Weekly зробіть висновок, чи наявні певні закономірності?

1.2. Використайте повний набір даних, щоб побудувати логістичну регресію, де Direction – залежна змінна, а п’ять зміщенних дохідностей та змінна Volume незалежні. Використайте функцію summary() для виводу результатів. Чи який-небудь з предикторів виявився статистично значущим? Якщо так, котрий?

1.3. Побудуйте матрицю помилок та обчисліть загальну частку правильних прогнозів. Поясніть, про що говорить матриця помилок щодо типів помилок, допущених логістичною регресією.

1.4. Тепер побудуйте модель логістичної регресії з використанням навчальних даних з 1990 по 2008 рр., з єдиним предиктором Lag2. Обчисліть матрицю помилок та загальну частку правильних прогнозів на тестових даних (тобто даних за 2009 та 2010 роки).

1.5. Повторіть 1.4, використовуючи лінійний дискримінантний аналіз.

1.6. Повторіть 1.4, використовуючи квадратичний дискримінантний аналіз.

1.7. Повторіть 1.4, використовуючи метод K-найближчих сусідів з K = 1.

1.8. Який із попередніх методів забезпечує найкращі результати?

1.9. Поекспериментуйте з різними комбінаціями предикторів, використовуючи у тому числі можливі перетворення та взаємодії. Наведіть відомості про змінні, метод та пов'язану матрицю помилок, що забезпечує найкращі результати. Зверніть увагу, що також варто поекспериментувати зі значеннями для K у класифікаторі K-найближчих сусідів.

2. У цій задачі розробимо модель, щоб передбачити, чи вибране авто має велике або низьке споживання газу на базі даних Auto.

2.1. Створіть двійкову змінну mpg01, яка містить 1, якщо mpg містить значення вище медіани, і 0, якщо mpg містить значення нижче його медіана. Ви можете обчислити медіану, використовуючи функцію median(). Може стати в нагоді функція data.frame () для створення єдиного набору даних, що містить як mpg01, так і інші змінні з Auto.

2.2. Дослідіть дані графічно, щоб виявити залежність між mpg01 та іншими характеристиками. Які характеристики здаються найбільш корисними для прогнозування mpg01? Розсіяні і коробчасті діаграми можуть бути корисними інструментами для відповіді на це запитання. Опишіть свої висновки.

2.3. Розбийте дані на навчальний та тестовий набори.

2.4. Застосуйте лінійний дискримінантний аналіз на навчальних даних, щоб передбачити mpg01 з використанням змінних, які виявилися найбільше пов'язані з mpg01 у 2.2. Яка тестова помилка отриманої моделі?

2.5. Застосуйте квадратичний дискримінантний аналіз на навчальних даних, щоб передбачити mpg01 з використанням змінних, які виявилися найбільше пов'язані з mpg01 у 2.2. Яка тестова помилка отриманої моделі?

2.6. Застосуйте логістичну регресію на навчальних даних, щоб передбачити mpg01 з використанням змінних, які, виявилися найбільше пов'язані з mpg01 у 2.2. Яка тестова помилка отриманої моделі?

2.7. Застосуйте метод К-найближчих сусідів з різними значеннями для К на навчальних даних, щоб передбачити mpg01 з використанням змінних, які, виявилися найбільше пов'язані з mpg01 у 2.2. Яка тестова помилка отриманої моделі? Яке значення К найкраще для даного завдання.

3. Дане завдання стосується написання функцій.

3.1. Напишіть функцію Power (), яка виведе результат піднесення 2 до 3-ї степені.

3.2. Напишіть нову функцію Power2 (), яка дозволяє передавати будь-які два числа, x і a, і виводить значення x^a.

3.3. Продемонструйте роботу функції Power2().

3.4. Напишіть нову функцію Power3 (), яка фактично повертає результат x^a як об'єкт R, а не просто друкує його на екран. Тобто, якщо ви зберігаєте значення x^a у об'єкті, який називається result у вашій функції, тоді ви можете просто return() цей результат, використовуючи такий рядок:

return (result )

3.5. Тепер, використовуючи функцію Power3 (), побудуйте графік f(x) = x2. Вісь *x* повинна відображати діапазон цілих чисел від 1 до 10 і вісь *y* повинна відображати x2. Позначте осі належним чином, і використовуйте відповідний заголовок для малюнка. Розгляньте можливість відображення будь-якої осі чи обох в логарифмічній шкалі. Це можна зробити використовуючи log = ‘‘ x ’’, log = ‘‘ y ’’ або log = ‘‘ xy ’’ як аргументи для функція plot ().

3.6. Напишіть функцію PlotPower(), яка дозволяє будувати графік функції x^a для фіксованого a та для діапазону значень x. Наприклад, ввівши PlotPower (1:10, 3), маємо отримати графік для значень х, що приймає значення 1, 2,. . . , 10 та вісь y, що приймає значення 13, 23,. . . , 103.

4. Використовуючи дані Boston пристосуйте модель класифікації для передбачення у вибраному районі рівня злочинності більшого чи меншого за медіану. Дослідіть логістичну регресію, лінійний та квадратичний дискримінантні аналізи, метод К-найближчих сусідів для різних значень К. Опишіть отримані результати.