PH227: AI and Data Science

Tutorial Sheet 3: Python/Data-Science Part

Note: Each tutorial sheet is worth 20 marks, and you have to submit the codes corresponding to the starred (*) problems to the TAs. The marks of the starred problems are indicated next to them.

1. * This problem deals with the numerical integration using trapezoidal and Simpson's rules. Suppose, we want to numerically compute the definite integral

$$I = \int_a^b f(x)dx.$$

For this purpose, we divide the domain of integration $[a, b]$ into n equal intervals of length $h = \frac{b-a}{n}$ so that we have $n+1$ evenly spaced points $(x_0, x_1, x_2, \ldots, x_n)$, with $x_0 = a$, $x_n = b$, and $x_{i+1} - x_i = h$, for $i = 0, 1 \ldots, n-1$.

- According to the trapezoidal rule, the approximate value of I is given by

$$I_{trap} = \frac{h}{2}\left(f_0 + 2f_1 + 2f_2 + \cdots + f_n\right) + O(h^2),$$

where $f_i = f(x_i)$, and in this formula the error scales as $\sim h^2$.

- According to the Simpson rule, if n is even, the approximate value of I is given by

$$I_{simp} = \frac{h}{3}\left(f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + \cdots + 2f_{n-2} + 4f_{n-1} + f_n\right) + O(h^4),$$

in this formula the error scales as $\sim h^4$. Note that this formula is valid only for even values of n, and for the same value of n, it is two orders of magnitude more accurate than the trapezoidal rule.

Write a single Python program which will ask the user for n, and then compute the following definite integrals using both the trapezoidal rule and the Simpson rule. Your program should reject odd values of n for the Simpson rule.

(a) $\int_0^{\pi/2} \sin x \, dx$ (exact value = 1)

(b) $\int_0^2 e^{-x} dx$ (exact value $= 1 - e^{-2} = 0.86466$ (correct up to 5 places)

(c) $\int_0^1 \frac{dx}{1+x^2}$ (exact value $= \pi/4$)

(d) $\int_1^2 \frac{dx}{x}$ (exact value $= \ln 2 = 0.693147$ (correct up to 5 places)

(e) $\int_0^{\pi/2} \sqrt{1 - \frac{1}{4}\sin^2 t}\, dt$ (exact value = 1.4675, correct up to 4 places)

(f) $\int_0^{\pi/2} \frac{dx}{\sin^2 x + \frac{1}{4}\cos^2 x}$ (exact value π) (10 marks)

2. * In this problem you have to make colored scatter plots from the famous Iris dataset listing the physical attributes and the species of 150 Iris flowers. This dataset is kept on the Moodle page of our course as the file iris.csv, and has five columns (attributes) consisting of sepal length, sepal width,

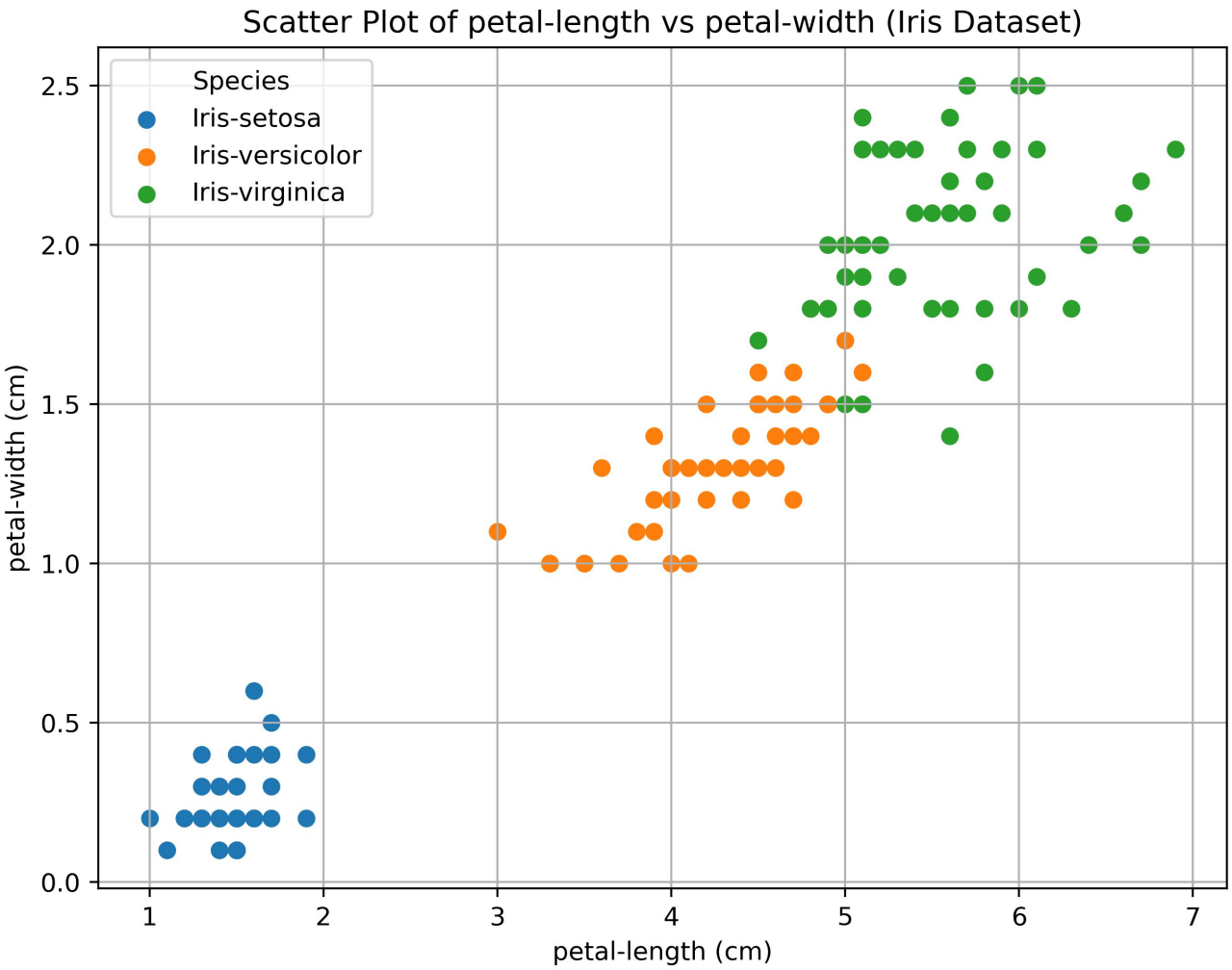

Figure 1: Petal length vs petal width scatter plot for the Iris data set.

petal length, petal width, and the name of the species of the Iris flower which can be Iris-setosa, Iris-versicolor, or Iris-viginica. The lengths and widths are in the units of cm. Write a Python program to make a colored scatter plot of this data using two user-chosen columns whose values will be used as (x, y), and the colors of the points will be determined by the species. For example, a scatter plot based on petal length and petal width is given in Fig. 1.