

## Assignment 1

---

### Instructions:

- There are 2 questions in this assignment, complete both.
  - Type your answers for both questions in a Word document. Indicate clearly the question number and part. Your submission should not exceed 11 pages.
  - Present your Python code for questions 1 and 2 in separate Jupyter notebooks, i.e. you should submit two Jupyter notebooks, one for each question. Indicate clearly which question number and part the code is for.
  - Submit the *Declaration of Academic Integrity* before submitting your assignment.
- 

### Question 1

(45 marks)

The United States Department of Agriculture (USDA) has its National Nutrient Database which compiles the nutritional information of food products. A subset of the data on dairy products is in the data file *dairy\_nutrition.csv*. The amount of nutrients per 100g of each dairy product is measured.

The following table lists the variables used in the file and their descriptions:

Variable	Description
Type	Type of dairy product: Yogurt, Cheese, Ice cream, Cream, Milk
Description	Description of the product
Protein	Amount of protein in grams
Fat	Amount of fat in grams
Carb	Amount of carbohydrates in grams
Sugar	Amount of sugar in grams
VitA	Amount of vitamin A in micrograms
Calcium	Amount of calcium in milligrams

- Should PCA be carried out on covariance or correlation matrix? Explain.
- Extract the principal components. Justify your decision and interpret the principal components. You should include the necessary tables, outputs and graphs.
- Which type(s) of dairy product has/have the following attributes? Explain your answer with the aid of a suitable graph with colour or marker to display ‘Type’ information.
  - Low carbohydrates and sugar but high in other nutrients.
  - High carbohydrates and sugar but low in other nutrients.

- (d) A dairy product has its nutritional value listed below. Which type of dairy product is it likely to be? Show your working and explain.

Protein: 4.8 g

Sugar: 19.1 g

Fat: 23 g

Vitamin A: 17 mcg

Carbohydrate: 26.2 g

Calcium: 121 mg

**Question 2**

(55 marks)

Glass is an often transparent solid material widely used for practical and decorative purposes in for example, windows and tableware. Glass of different chemical compositions and refractive properties may be used for different type of objects. Refer to the data file *glass.csv*. There were 102 objects generally classified into 5 classes: building window (BuildWin), vehicle window (VehWin), tableware, containers, and headlamps. The percentage of the corresponding oxide (by weight) for the elements sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and refractive index (RI) of each object are recorded.

- (a) Should PCA be carried out on covariance or correlation matrix? Explain.
- (b) Extract the principal components. Justify your decision and interpret the principal components. You should include the necessary tables, outputs and graphs.
- (c) The following shows the attributes of a glass object. Which class does it likely belong to? Explain your answer with the aid of a suitable graph with colour or marker to display 'Class' information.

RI: 1.51641

Si: 73.05

Na: 13.04

K: 0.53

Mg: 3.5

Ca: 8.6

Al: 1.28

- (d) Explain how PC3 is advantageous over the first two principal components.