MS0240 Mathematics for AI

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 10 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 (50 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 food products is in the data file *food_nutrition.csv*. The amount of nutrients per 100g of each food product is measured.

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

Variable	**Description**
Type	Type of food product: Breakfast Cereals, Fruits and Fruit Juices, Poultry Products, Vegetables and Vegetable Products
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
Fiber	Amount of fiber in grams
VitA	Amount of vitamin A in micrograms

(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) Which type(s) of food product has/have the following attributes? Explain your answer with the aid of a suitable graph with colour or marker to display '*Type*' information.

(i) Low protein and fat but high in other nutrients.

(ii) High protein and fat but low in other nutrients.

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

Protein: 0.3 g
Fat: 0.1 g
Carbohydrate: 19.6 g

Sugar: 16.7 g
Fiber: 2.9 g
Vitamin A: 2 mcg

Question 2 (50 marks)

Wines grown in the same region in Italy but by three different cultivators were analysed. The analysis determined the quantities of 7 constituents found in each wine. Results of the chemical analysis can be found in the file *wine.csv*. The variable "cultivator" indicates which of the three cultivators each wine originated. The remaining variables are names of each constituent.

(a) Should PCA be carried out on covariance or correlation matrix? Explain.

(b) Suppose we will retain the principal components with cumulative percentage of at least 80% of the total variance, extract the principal components and interpret your results. You should include the necessary tables, outputs and graphs to justify your answer.

(c) The following shows the attributes of a particular wine. Which cultivator is it likely to originate from? Explain your answer with the aid of a suitable graph with colour or marker to display '*cultivator*' information.

alcohol: 12.75 %v/v at 20°C

malic acid: 3.11 g/L

magnesium: 96 mg/L

flavanoids: 0.61 mg/L

nonflavanoid_phenols: 0.52 mg/L

color_intensity: 4.56 units

proline: 632 mg/L

(d) Discuss whether PC3 is useful in helping you identify the cultivator the wine in part (c) is likely to originate from.
