MS0240 Mathematics for AI

Assignment 1

Instructions:

- There are 2 questions in this assignment, complete both.
- Based on the **last digit of your admission number**, download the corresponding datasets on dairy nutrition and cars. For example, students with admission number Pxxxxxx6 should use dairy_nutrition_6.xlsx and cars_6.xlsx. Marks will be deducted for the wrong datasets used.
- Perform Principal Component Analysis on both datasets in order to answer the questions pertaining to each dataset.
- Present your analysis in a Word document. Indicate clearly the question number. Your submission should not exceed 12 pages.
- Present your Python code for datasets 1 and 2 in separate Jupyter notebooks, i.e. you should submit two Jupyter notebooks, one for each dataset. Indicate clearly which question number the code is for.
- **All tables, outputs, graphs, calculation results used for analysis and explanation should be presented in the document.** The tutor only refers to your Python codes to see how you arrive at your outputs and does not look for the outputs in the Jupyter notebooks.
- An oral presentation may be required at your tutor's discretion.
- Submit the *Declaration of Academic Integrity (A1)* before submitting your assignment.
- Use the following template to acknowledge the use of any AI Tools.

Name of AI tool	< For example, ChatGPT >
Input prompt	< Insert the question that you asked ChatGPT >
Date generated	< Insert the date that ChatGPT response was generated, since ChatGPT is an evolving technology >
Output generated	< Insert the response verbatim from ChatGPT >
Impact on submission	< Briefly explain which part of your submitted work was ChatGPT's response applied >

MS0240 Mathematics for AI

Question 1

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_*(digit)*.xlsx*. 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: Cheese, Cream, Ice cream, Milk, Yogurt
Description	Description of the product
Protein_g	Amount of protein in grams
Fat_g	Amount of fat in grams
Carb_g	Amount of carbohydrates in grams
Sugar_g	Amount of sugar in grams
VitA_mcg	Amount of vitamin A in micrograms
VitB6_mg	Amount of vitamin B6 in milligrams
VitB12_mg	Amount of vitamin B12 in milligrams
Calcium	Amount of calcium in milligrams

(a) Present your PCA analysis with all the necessary outputs and graphs. Explain all decisions made in the analysis.

(b) Which type(s) of dairy product has/have the following attributes?

(i) Low carbohydrates and sugar but high in other nutrients.

(ii) High carbohydrates and sugar but low in other nutrients.

(c) 2 dairy products have their nutritional values listed below. Which type of dairy product is each of them likely to be? Use a suitable number of principal components to help you with your analysis.

Product 1:

Protein: 22.17 g
Fat: 22.35 g
Carbohydrate: 2.22 g
Sugar: 1.01 g
Vitamin A: 181 mcg
Vitamin B6: 0.034 mg
Vitamin B12: 2.28 mg
Calcium: 505 mg

Product 2:

Protein: 4.32 g
Fat: 1.42 g
Carbohydrate: 23.0 g
Sugar: 14.58 g
Vitamin A: 13 mcg
Vitamin B6: 0.047 mg
Vitamin B12: 0.53 mg
Calcium: 114 mg

(d) Describe your observations so far, comparing what you have done in part (c) and the decision(s) you have made in the earlier part of PCA in (a). Compare how you may have expected the principal components to perform and how they have actually performed.

Question 2

A car magazine published in 1985 compiled the attributes of 195 cars. The data can be found in the file *cars_*(digit)*.xlsx*.

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

Variable	Description
Brand	Brand and model of vehicle
Type	Type of vehicle: Sports Car (Sports), Sports Utility Vehicle (SUV), Wagon, Minivan, Sedan
Wheel-base	Wheel base (inches)
Length	Length (inches)
Width	Width (inches)
Height	Height (inches)
Curb-weight	Weight (Pounds)
Cylinders	Number of cylinders
Engine	Engine size (litres)
Compression-ratio	Compression ratio
Horsepower	Horsepower of vehicle
City-mpg	City Miles Per Gallon
Highway-mpg	Highway Miles Per Gallon
Price	Price (US$)

(a) Present your PCA analysis with all the necessary outputs and graphs. Explain all decisions made in the analysis.

(b) Explain the difference between the PCA results of this dataset and the dairy nutrition dataset in Question 1, and thus comment on the usefulness of PCA for classification or clustering purposes.
