Name: _____

Campus ID: _____

Total Marks: 10
Time Duration: 15 minutes

Question 1 (5 marks)

For each of the following scenarios, (a) determine whether the task is supervised, semi-supervised, or unsupervised. (b) Also identify whether it is a regression, classification, or clustering problem.

(a) You have a dataset containing house prices and corresponding features like square footage, number of bedrooms, and location, and your goal is to predict the price of a new house.

(b) You are given a set of customer reviews, and you want to group them into positive, neutral, and negative categories based on the sentiment expressed.

(c) You are working with a dataset that contains sales figures for different products over time, and you need to predict the sales for the upcoming months.

(d) You have a large number of unlabeled images of handwritten digits, and your task is to group similar images together to identify patterns.

(e) You are given a dataset of patient records, with some labeled as having a disease and others not, along with a larger set of unlabeled records. You want to predict whether the patients in the unlabeled dataset have the disease.

(f) You are analyzing a dataset containing information on people's income and education levels, and your goal is to classify them into different income brackets (low, medium, high).

(g) You have data on weather conditions, including temperature and humidity, and your goal is to identify patterns and categorize days with similar weather characteristics.

(h) You are given a dataset with labeled photos of cars and trucks, and your task is to create a model that can identify whether a new image contains a car or a truck.

(i) You have a dataset that contains ages and corresponding heights of a group of individuals, and your goal is to predict the height of a person based on their age.

(j) You are provided with a collection of customer transactions, with no labels, and you need to segment customers based on their purchasing behavior.

Question 2 (5 marks)

Consider the following list of applications. For each application, identify whether it commonly involves the use of sparse vectors. Provide a brief justification to support your answer.

(a) A recommendation system that suggests products to users based on their browsing history and purchase patterns. [1 mark]

(b) An image processing algorithm that applies filters to a large collection of high-resolution images. [1 mark]

(c) A document retrieval system that indexes and searches through a vast number of text documents to find relevant ones based on keyword queries. [1 mark]

(d) A weather forecasting model that uses historical temperature data, atmospheric pressure, and humidity to predict future weather conditions. [1 mark]

(e) A gene expression analysis where a matrix of gene expression levels is used to identify patterns across different samples, with many genes not being expressed in every sample. [1 mark]