Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Homework 1

Show your work include any code snippets that you used to generate answers. Complete this assignment individually.

1. What are the two main types of attributes typically find in data. [2 points]
2. Consider the following matrix D; and answer all the questions. [14 points]
   1. What is the sample mean of this data attribute ? [2 points]
   2. What is the sample covariance between attributes and ? [2 points]
   3. What is the sample multi-dimensional mean of this data matrix? (Your answer should be a vector) [2 points]
   4. What is the sample covariance of the attribute [2 points]
   5. What is the covariance matrix of D? [2 points]
   6. What is the correlation coefficient between the attributes and ? [2 points]
   7. What is the total variance of this matrix D? [2 points]
3. Consider the following 5-dimensional vectors: [6 points]
   1. What is the [2 points]
   2. What is the [2 points]
   3. What is the angle between the vectors and ? [2 points]
4. Consider the following matrix D.
   1. Use the One-Hot encoding method to transform the categorical data into numerical data in the following matrix. You can assume that the attribute can only contain 4 values: [2 points]
   2. What is the value of (based on the transformed data matrix)? [2 points]
5. The following questions reference the Heart Disease data set from the UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>  
   Answer the following questions about the dataset.
   1. How many rows (entities/instances) are there in this dataset? [1 point]
   2. How many attributes are in this dataset? [1 point]
   3. What is kind of data is stored in the attribute “cigs”? [1 point]