Yurii Oleksandr Bibik

Dr. Kang

CS 422

15 February 2023

Homework 1

Task 1: Visualize the MNIST data using PCA. Reduce the data dimension to two or three and plot the data of reduced dimension. Must plot all the data of ten groups (0 to 9). (40 points)  
 In the code, I firstly imported the necessary libraries. Then, I read the MNIST data from a csv file and separated the target variable (the digit labels) and the feature set (the pixel values). I used PCA (Principal Component Analysis) from the scikit-learn library to reduce the dimensionality of the feature set from 784 to 2. Next, I visualized the transformed data using a scatter plot, where each group of 100 data points (representing a single digit) is plotted with a different color. Then I added labels to the x, y axis and the chart itself. I also added a legend to the plot to show which color represents which digit. Finally, I displayed the plot using plt.show().

Task 2: Visualize the columns of K, M, and N in the housing data using Boxplot. (30 points)

In the code I used the pandas and matplotlib libraries to visualize the columns "K", "M", and "N" from the housing dataset. Firstly, it reads the data from a .csv file. Then I renamed the columns from their numbers to "K", "M", and "N". Then, I created a box plot of the data using the matplotlib library. Then I added labels to the x, y axis and the chart itself. The box plot provides information about the density and distribution of the data values at different values. Finally I displayed the boxplot using the plt.show() command.

Task 3: Visualize the column of A in the housing data using histogram. (30 points)  
 In the code I visualize the distribution of the data in column 0 (referred to as "A") of a housing dataset. Firstly, I import the necessary libraries. Then I read the housing data from a CSV file. Then I create a histogram of the data in column 0 (“A”). I added labels to the x, y axis and the chart itself. Finally, I displayed the plot using plt.show().