CS 165B – Machine Learning, Summer 2020

Homework 4

Due *Thursday, September 10, 2020 11:59pm*

Write a Python3 program called `hw4.py` that implements principal component analysis (PCA) and perform dimensionality reduction with PCA on the provided dataset.

1 Details

The program should be able to read in the contents of files and perform dimensionality reduction with PCA. You are provided a skeleton code for this assignment. Avoid changing the signature of the provided functions. In addition to implementing the algorithm, you also need to generate a scatterplot of the first and second principal component of the provided dataset after dimensionality reduction. The image of the plot needs to be submitted along with the code.

The program must be able to run on CSIL and it must finish running within 1 minute. You are not allowed to use any third-party libraries or frameworks (not including the standard library) for this homework except those declared in the skeleton code. Be aware that NumPy, one of the allowed libraries, provides methods for computing the covariance matrix, eigenvalues and eigenvectors, which may be useful for this homework.

2 Grading

Your submission will be graded on both the correct implementation of PCA as well as the plot. The code must work on CSIL or you may receive 0%.

Grade Breakdown:

- 40% for correct PCA (i.e. returns the correct eigenvectors and eigenvalues)
- 40% for correctly transforming the dataset
- 30% for plotting the dataset using the first and second principal component

3 Submission

Submit *only* `hw4.py` and the plot to GauchoSpace. Submit them directly. In other words, do not submit them in a folder and do not zip the files.