Cem Ergin

COMP 421 HW#6 Report:

I’ve read the chapter concerning LDA, re-watched the classes and even read sources online but for some reason, I couldn’t get my functions to provide a meaningful dimensionality reduction.

In my 2-D scatters plots, most of the time I observed the data and data from all classes seemed as if scattered randomly around a common origin which is the exact opposite of what LDA should provide. Due to this problem, my 5-nearest neighbor classifier couldn’t provide an accuracy rate larger than 0.10 which suggest that I was really successful in scattering the data points randomly while trying to reduce the dimensionality. After some work, the randomness of the training data on the 2-D plot improved drastically but I didn’t have much luck with the test data.

I think that this happened due to the inner workings of numpy I’m not aware of and due to my insistence of trying to use sets to hold the unique class labels. I think that the iterator of python doesn’t output these labels according to the order in Label data and therefore corrupting everything I’m trying to do. This could have been easily prevented if I didn’t start on the wrong foot. Will be more careful next time.

To showcase that as least the 5-nearest neighbor classifier is working I added an initial plot of something I call “The Donut Case” in which there are 2 different classes and when plotted they resemble a donut. I used those data to prove that my classifier is actually working pretty successfully; it outputs an accuracy between 0.95 – 1.00 when used on the donut case. Additionally, my KNN algorithm needs “ sortedcontainers” package to work which was a simple library that helps you create sorted lists, sets, etc. as the name implies.