Assignment 1: PCA vs. LDA
Taten Knight

I had very little experience with the concepts of PCA and LDA prior to reading this article, so I spent some time on the StatQuest YouTube channel prior to reading the article in order to gain a basic understanding of the two methods. I may reference those videos as well as the article.

The goal of this paper on the whole was to test the intuition that Principal Component Analysis (PCA) can outperform Linear Discriminant Analysis (LDA) in certain situations. Principal component analysis is a method of finding transformed dimensions along which the data has the maximum variance, and ranking those dimensions in a way that shows us which Principal Component (PC) has the greatest effect on the variance. LDA is a method that is used to find the maximum variation between known categories of data. I'd like to try to put this in my own words. PCA creates new axes along which the data is the farthest away from the origin on average, while LDA tries to find the axes that are the most effective at separating the data.

The study took an extremely small sample size of faces to show that PCA can yield better result when only a small bit of training data is available, while with increased sample size, LDA method quickly proves itself to be superior. The task at hand was the recognition of which individual a face image belonged to. It should be noted and re-noted that the purpose of LDA is to find that best way to segment known groups of data, while PCA is only trying to find the components that best describe the data. It is no surprise that the components that because describe the data or a group of data does not necessarily provide the best way to segment that subsequentially group new images.

Although on large data sets LDA will almost always perform better, there are small sample situations where PCA may yield better results for categorization.

In addition to relative performance of the algorithms, I also learned that neglecting the first 3 eigenvectors found via PCA tends to increase performance because these three PCs tend to correlate highly to the intensity data of the facial images and not the face itself.

It was a cool opportunity to read an older paper about PCA vs. LDA and compare what I knew about modern approaches with those from 20 years past. It seems that LDA is still the inferior method for categorization due to focusing on variation in all the data vs. variance between known data points, but the paper also mentions that transformation into a PCA space as a potential intermediate step for performing LDA, showing that the tools can work in tandem to better solve the problem than they may alone. I need to take the time to dive into some more modern literature on the topics to see where we are now and how these approaches have evolved.