# Bayes’ and k-Nearest Neighbor Classification

Using Original, PCA, and LDA Datasets

Michael Scuderi
ENME 439M - Project 1
University of Maryland - College Park

# Table of Contents

# Introduction

For my project I decided to use the ‘face’ dataset. This dataset contains faces of three different classes: neutral expression, non-neutral facial expression, and an illumination variant of one of the other faces. The main reason I chose to use this dataset was due to its format. The first through third image of the dataset was of class 1, 2, 3, respectively and kept that pattern throughout. This made it simple to divide the dataset equally into testing and training sets. The ‘face’ dataset was organized as a 600x24x21 array, which I then flattened to 600x504 for the training and testing of the model. Flattening the data made it easy to perform statistical operations on the data.

After formatting the image data into a 600x504 array, I then trained and tested the model for Bayes’ Classifier and k-Nearest Neighbors. Training and testing was then done with the original, LDA, and PCA datasets.

# Statistical Methods

## Original Dataset

For the first experiment, the only preprocessing to the data that was done was dividing the data into training and testing sets. The training test was then used to calculate mean and variance for each class using maximum-likelihood estimation. Since the original data is being used, the main parameter that can be changed and tested is the training size. Training sizes from 99 to 507.Training sizes were chosen such that each class had an equal amount of images in the training set. This allowed the decision to simplify to maximizing likelihood.

### Bayes’ Classifier

For Bayes’ Classifier, using a training set of roughly 50% of the dataset yielded the best classification success rate of 90.9%. Smaller training sets’ lower success rate can be attributed to MLE not being accurate to the underlying distribution of the data due to the limited samples. Larger training sets lose some accuracy due to possible overfitting. In addition, a smaller training set is more practical due to increasing computation and amounts of data required.

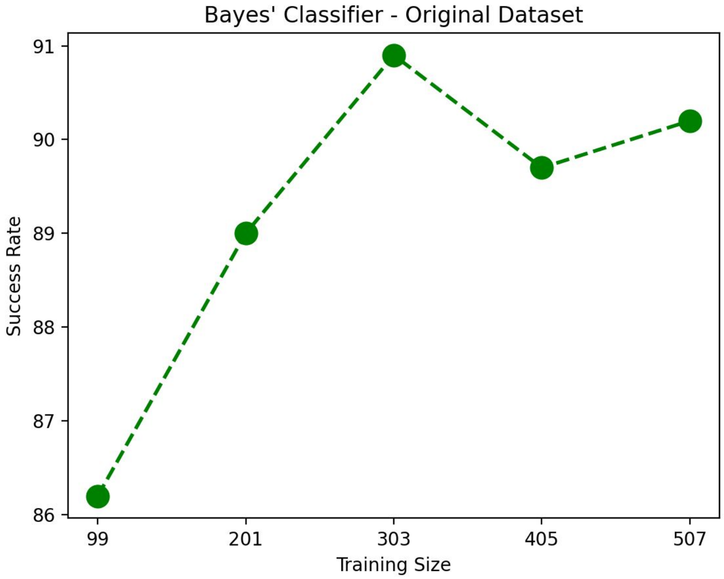

Figure 1: Bayes’ Classifier success rate vs. training size

## k-Nearest Neighbors

For k-Nearest Neighbors, I performed the same analysis of the training set along with changing k values. Amongst all k values, a larger training set seemed to make the classifier perform worse. A k value of 5 performed the best out of the k values tested. It makes sense that k=1 performed the worst because there is a higher probability of misclassification due to only considering 1 neighbor; a neighbor that strays into another class region can cause misclassification of x’. I expected classifier success to increase as training size increased, however, the opposite was true in my experiment. I think this may have happened due to many features (pixels) in the images having similar values in certain regions such as the head and nose, which differ very little between classes. There may be very little space between class means for certain features, making the classifier prone to misclassification.

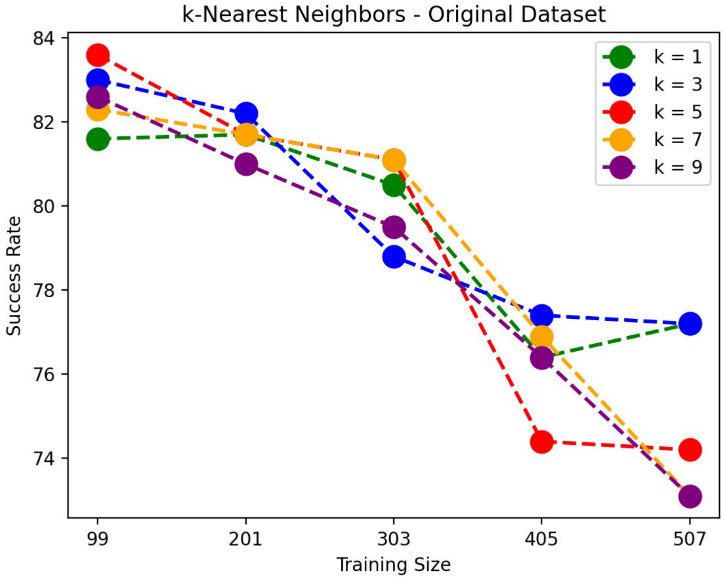

Figure 2: k-Nearest Neighbor success rate vs. training size vs. k value

# Principal Component Analysis

When applying principal component analysis (PCA) to the dataset, I was able to determine that 90% of variance could be represented with 21 features and 95% of variance could be represented with 61 features. It is also important to note that principal component 1 had a very large portion of variance in the images covered; over 70%. I then carried out the experiments comparing performance of the classifiers while varying variance and training size.

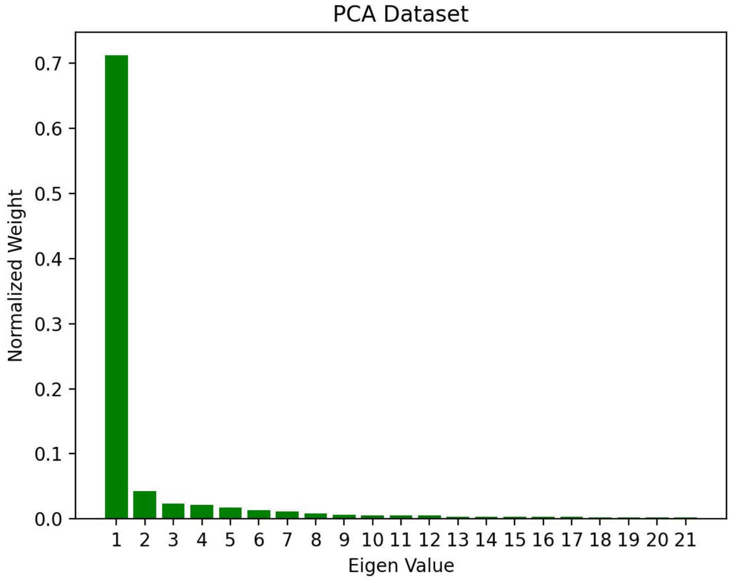

Figure 3: Principal components and their weights with respect to sample variance represented

## Bayes' Classifier

Testing different training set sizes vs. different variance coverage from PCA, I was able to identify Bayes' Classifier success rate peaks at 95% variance coverage with a training set size of 507 images. From the graph below, it is evident that the success rate of Bayes' at 90% variance peaks at the training size of 405 and then dips down. This could be evidence that the model begins to overfit at this point. However, with 95% variance, the model does not overfit as quickly and is able to reach peak performance with a training size of 507. This makes sense because the 95% variance model has more features which could increase the amount of data required to overfit, depending on the features' distributions. Overall, the Bayes' Classifier did not perform to my expectations for this test, as the success rate is rather low.

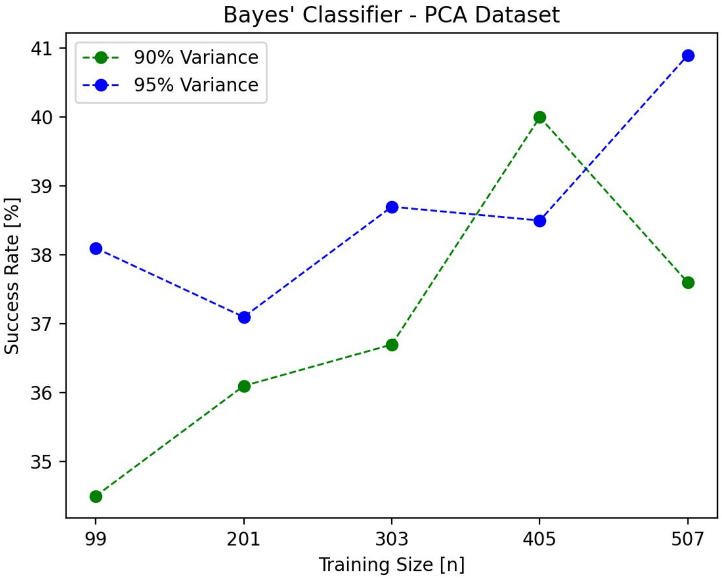

Figure 4: PCA results for Bayes’ Classifier

## k-Nearest Neighbors

Since k=5 was optimal in the original dataset, the following test was performed with a constant k value of 5. When testing the kNN classifier, a local maximum is found with a training size of 201 and 507. Evident from the below graph, variance coverage by PCA did not have an affect on this dataset’s success rate. 90% and 95% variance tests both produced the same performance in classifying the images.The kNN algorithm with PCA applied data had a low success rate, similar to Bayes’.

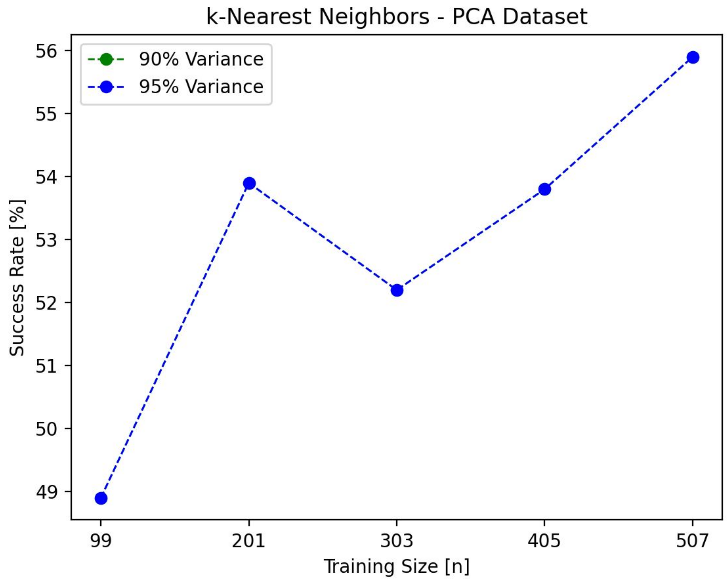

Figure 5: PCA results for k-Nearest Neighbors

# Fisher’s Linear Discriminant Analysis

Applying LDA to the dataset, it was apparent that one feature held almost all variance in the images. The rest of the features contributed very little to variance in comparison to eigenvalue 1, however, their contribution to training and testing still made a noticeable difference.

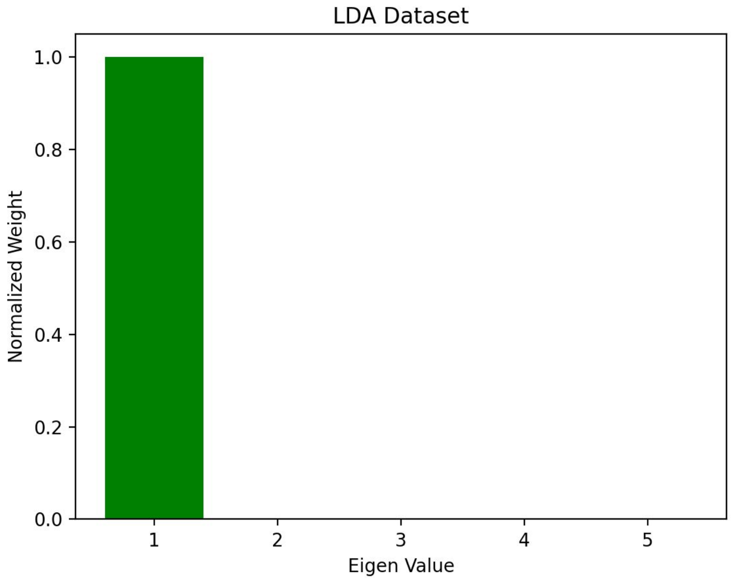

Figure 6: LDA eigenvalues and their weights with respect to each other

## Bayes’ Classifier

From the figure below, we can determine a smaller training size is beneficial for the Bayes Classifier using LDA.

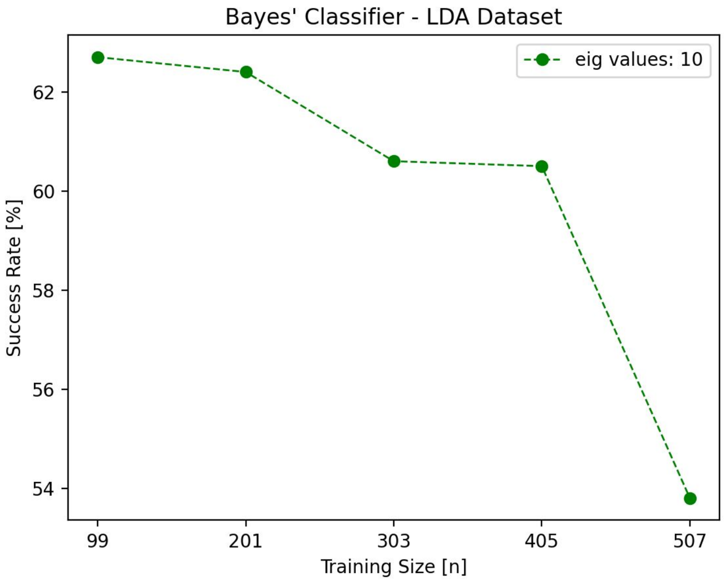

Figure 7: LDA training set size comparison

Moving forward, multiple quantities of eigenvalues were tested. Bayes’ Classifier using the top 50 eigenvalues and eigenvectors produces the best return on success vs. compute time. It is not far off from the original dataset success rate, but only roughly 1/10th of the calculations is required.

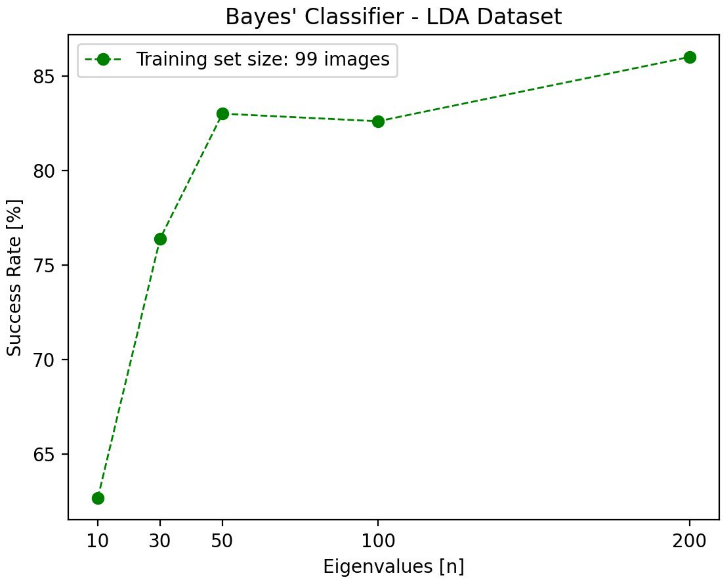

Figure 8: LDA results for Bayes' Classifier

## k-Nearest Neighbors

For the LDA kNN algorithm, performance had a local peak with a training set of 405 images. A training set of 405 images was used moving forward with the eigenvalue experiments.

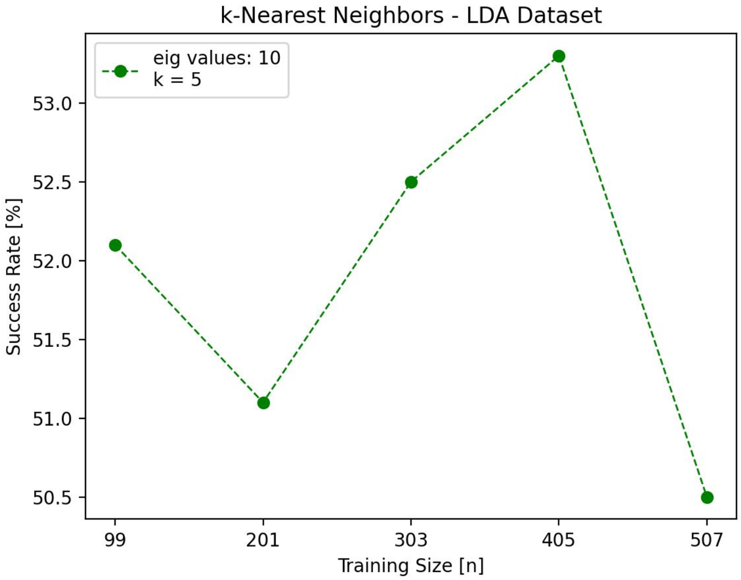

Figure 9: LDA training set size comparison for kNN

After testing many different quantities of eigenvalues, it is clear that kNN benefits greatly from more eigenvalues, more so than Bayes’ Classifier.

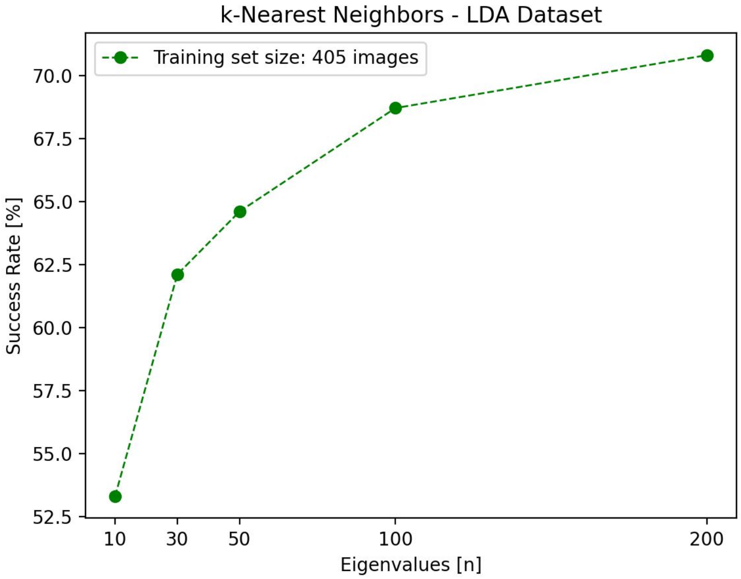

Figure 10: LDA kNN results

# Results

Overall, the best performing classifier was the Bayes’ Classifier using the original dataset with a success rate of 90.9%. However, due to the compute time needed to use every feature in the image, LDA might be the most beneficial depending on the application of the classifier due to only having to compute 1/10th of the data. Similarly, k-Nearest Neighbors sees the best performance with the original dataset, but achieves satisfactory success using LDA (70% vs 83%). the k-Nearest Neighbors algorithm took a lot longer to test than the Bayes’ Classifier. However, this could probably be optimized in the software engineering phase of the classifier.