

Facial Classifier

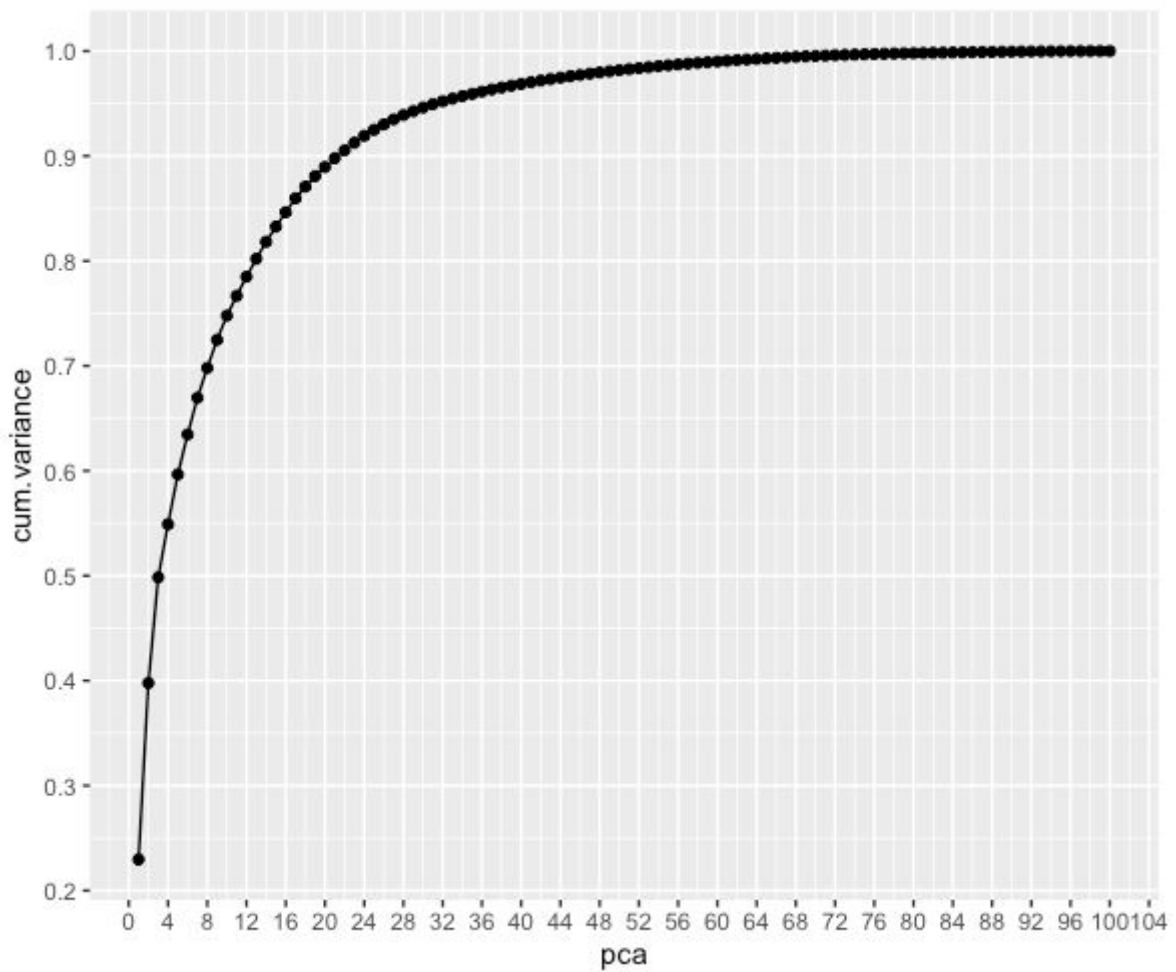
Subject: Statistical Learning

Name: Stanisław Czembor

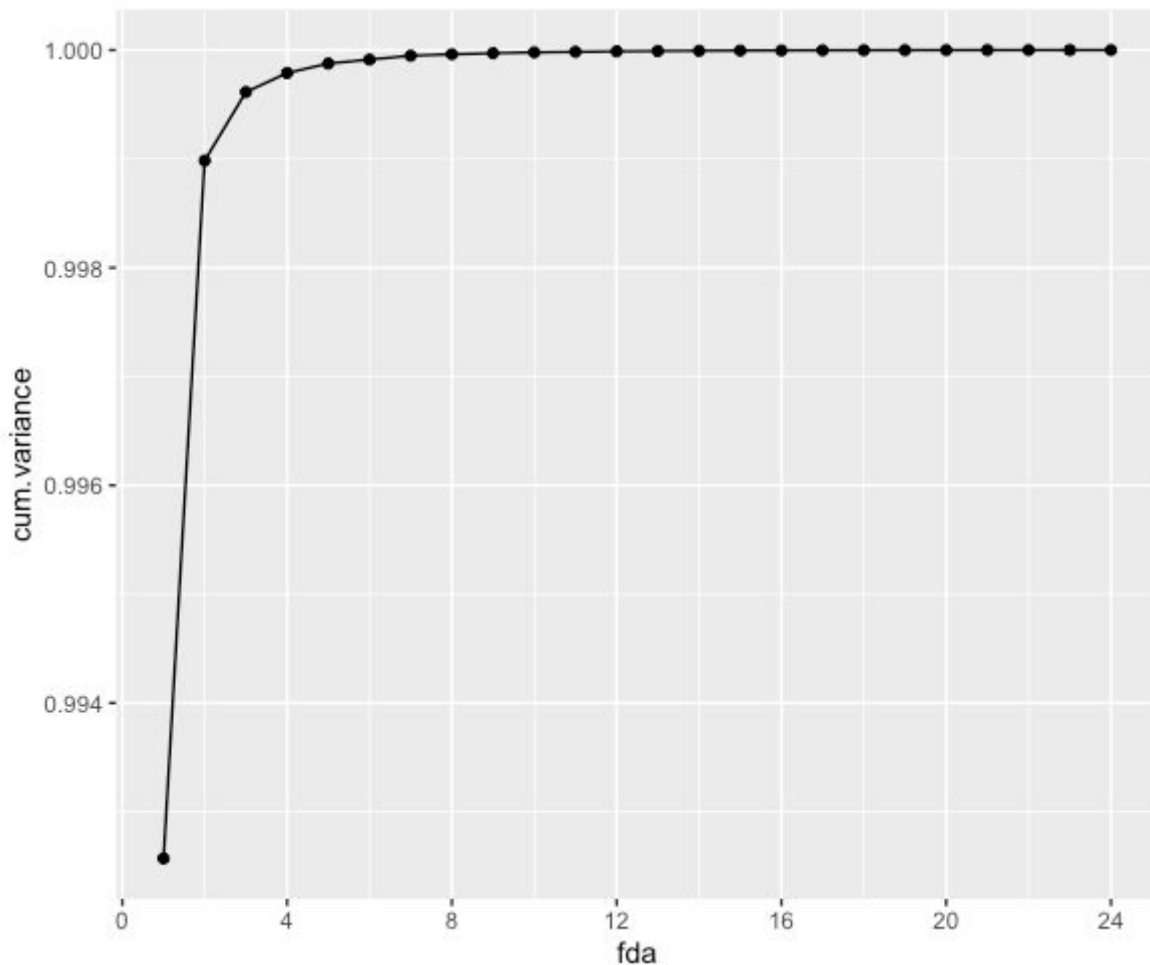
Group: 96

Date: 04/12/20

1. After running the pca function on the training data set I retrieve the eigenvalues from the function and compute the cumulative sum of the variance. The plot below shows the obtained results. I decided to use 24 pca (number of classes -1). The cumulative sum of variance for it is: **0.9175214**



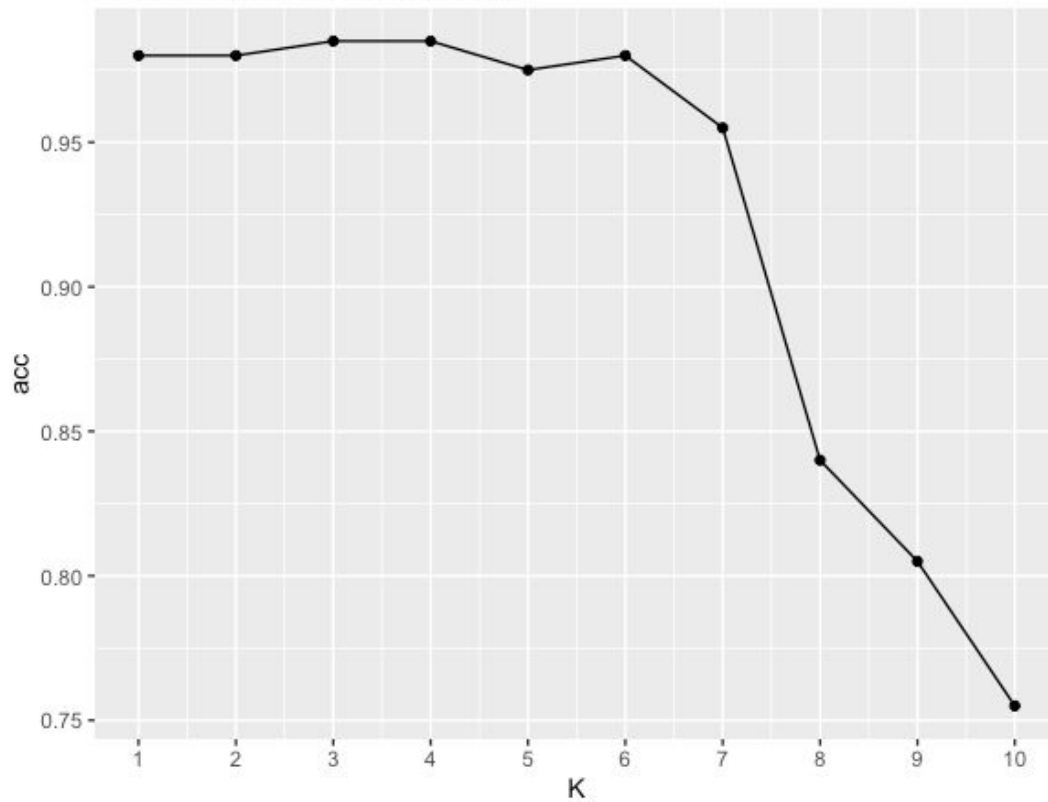
2. After scaling and projecting the training data set I run FDA on it. I obtained the following results. After analyzing the plot below I decided to go with 8 first eigenvectors. The cumulative sum is equal to: **0.9973299**.



3. The plot below shows the results (accuracy) after running repeated stratified random sampling, at the same time changing the number of k-nearest neighbors. I measured the performance for two different similarity metrics: Euclidean and Manhattan. Each dot on the plot responds to the mean of 4 results for given k. Looking at the results I decided to use euclidean distance and k=4, since it accomplished the highest accuracy. Important thing to note is that after every sampling I would run pca and fda only for the training set and then project the test set with the obtained eigenvectors. Running pca for the entire set is not a good practice, we shall not see any test data, and shall not use it to obtain any parameters for the learning algorithm.

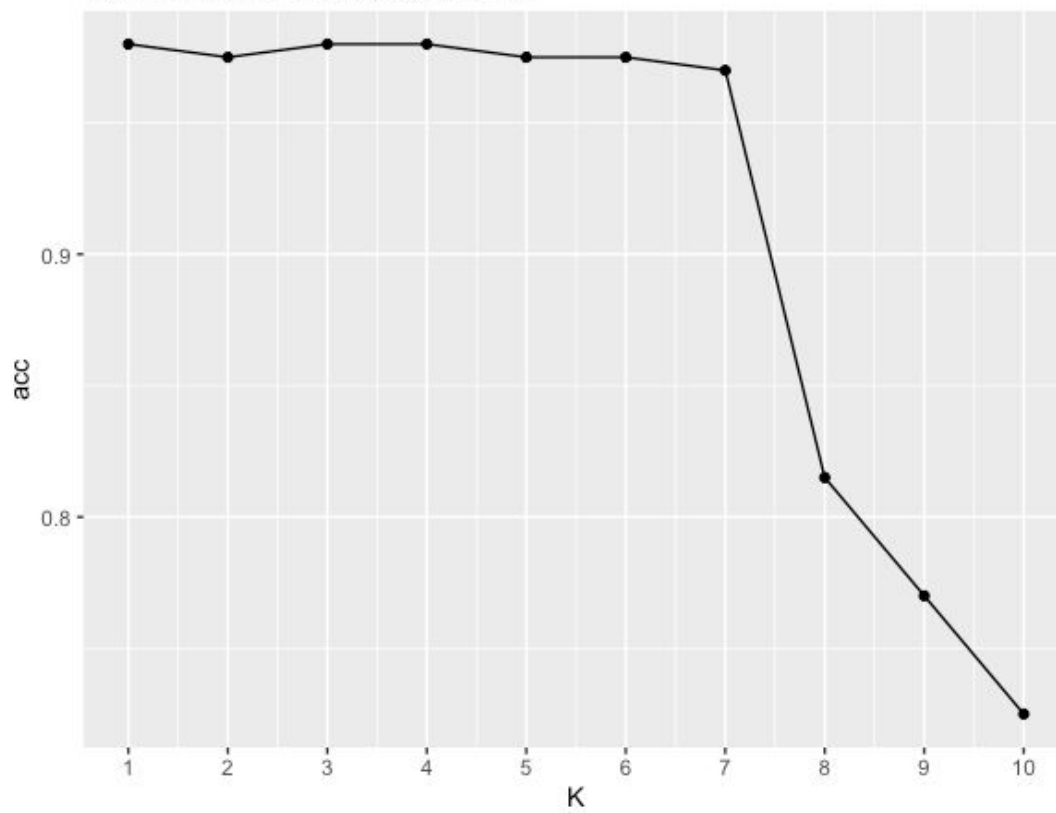
Accuracy for given K nearest neighbours

Similarity metric: Euclidan Distance



Accuracy for given K nearest neighbours

Similarity metric: Manhattan Distance



3. While looking for the optimal threshold I run a loop trying different denominators for the threshold function I wrote. I chose the value of `threshold_value[32]`. choosing higher values will result in overfitting(All the pictures will be classified as 0). The following plot below shows the obtained results. **`threshold_value[32] = 1400.012`**

