

Ryan Farr – rfl238
October 21st, 2018
CS 5785 – Applied Machine Learning

Homework 2

1. Eigenface for face recognition
 - a. The face dataset was downloaded
 - b. Below is the 10th training data image



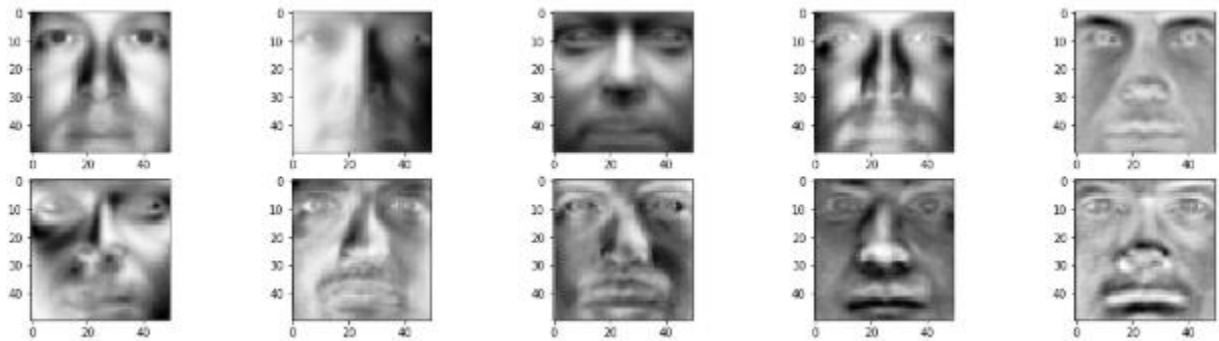
- c. Below is the average face



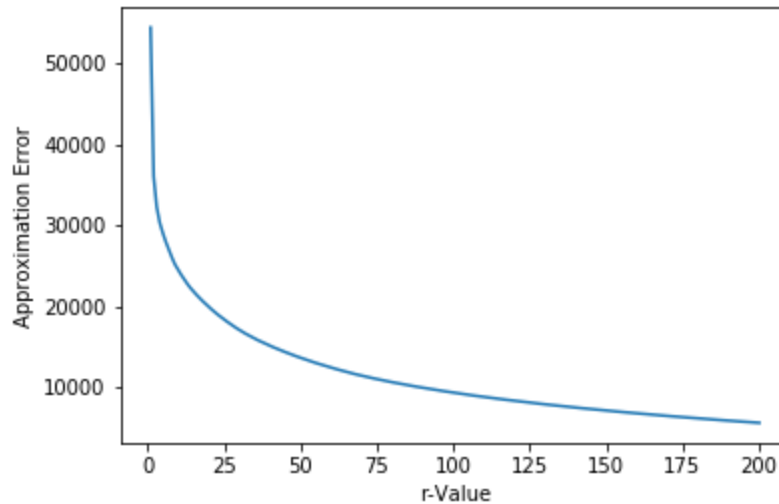
- d. Below is a training face minus the average face (left) and a testing face minus the average face (right).



e. Below are the first 10 eigenfaces.

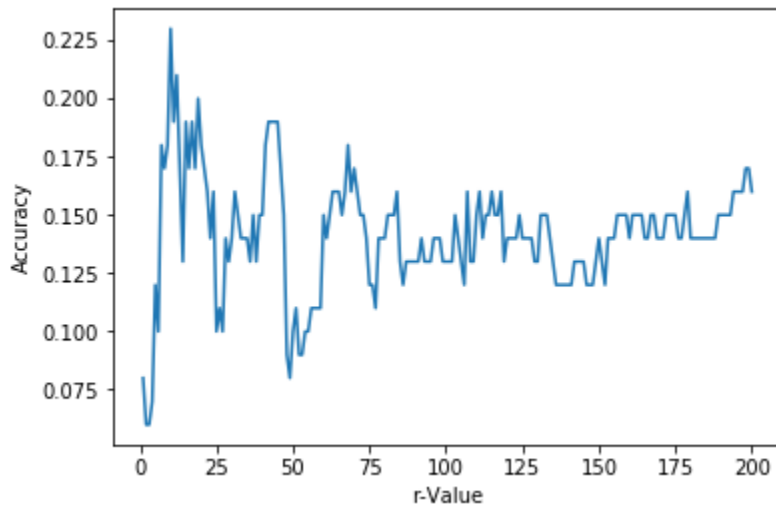


f. Below is a graph of the rank-r approximation error with values of r between 1 and 200. As expected, with higher values of r we're getting lower error and a generally better approximation of f .



g. Completed. See “Part 1.G” in main.ipynb

- h. Below is a graph of the accuracy of a logistic regression classifier using values of r between 1 and 200. As can be seen, the accuracy actually tends to decrease and stabilize as we go towards $r=200$. This is interesting because our intuition might tell us that the accuracy should increase as we get a better approximation of X , but we also need to note that we’re adding noise.



2. What’s cooking?

a. Completed, username is rlf238.

b. The number of training samples is 39774, the number of categories is 20, and the number of unique ingredients is 7,137.

c. Completed, see Part 2.C in main.ipynb

- d. Gaussian 3-fold cross validation accuracy: 0.3794, Bernoulli distribution 3-fold cross validation accuracy: 0.6784.
- e. In this case, the Bernoulli prior assumption gave a much better score. This is due to the fact that the data itself is 0 or 1 as opposed to floating point gaussian, where the Gaussian assumption likely would have done better. That is, the format of the data highly influenced which assumption is more correct.
- f. Using logistic regression the accuracy on 3-fold cross validation is 0.7757, higher even than Naïve Bayes with the Bernoulli prior assumption.
- g. Training logistic regression on the entire dataset, testing on the testing set and reporting the results to Kaggle produced accuracy of 0.78338.

Your most recent submission				
Name	Submitted	Wait time	Execution time	Score
testing_output.csv	just now	0 seconds	0 seconds	0.78338
Complete				
Jump to your position on the leaderboard ▼				