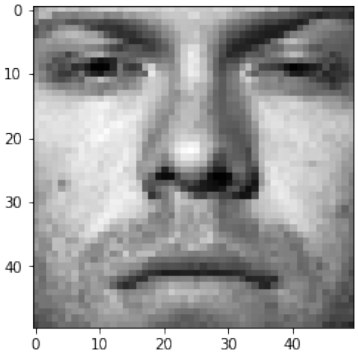
CS 5785: Homework 1 Write-up

Vianne Gao
vrg36@cornell.edu

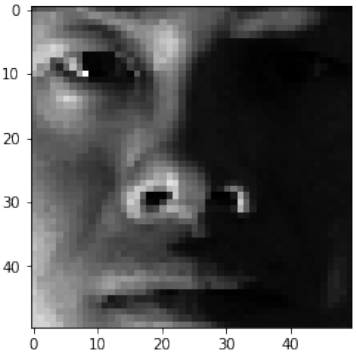
22nd October 2018

1. Eigen Faces

(b) Display a face image from training data and test data.

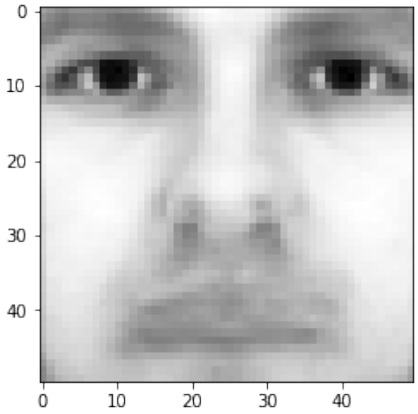


Sample train.



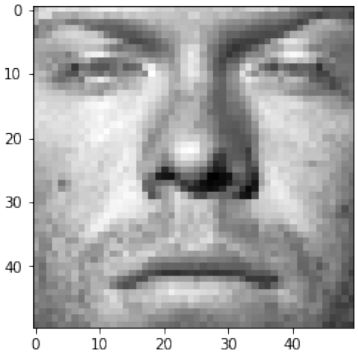
Sample test.

(c) Display the average face.

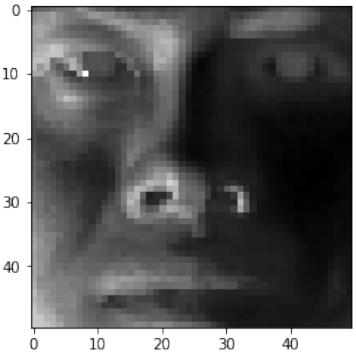


Average face.

(d) Mean subtraction.

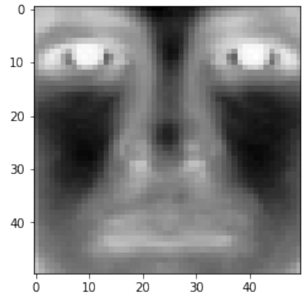


Sample train.

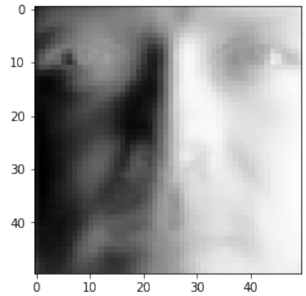


Sample test.

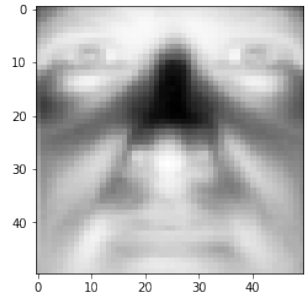
(e) Display the first 10 eigenfaces.



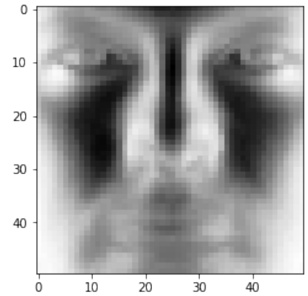
Sample train.



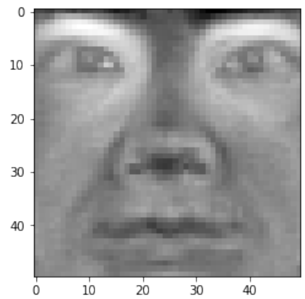
Sample test.



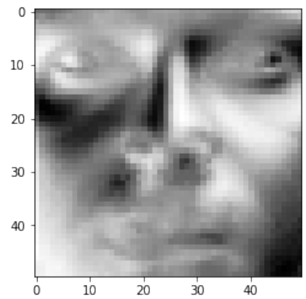
Sample train.



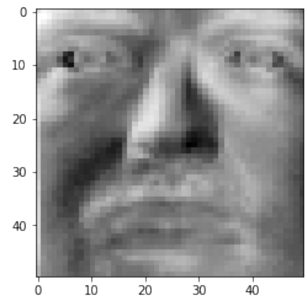
Sample test.



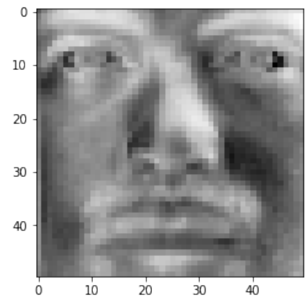
Sample train.



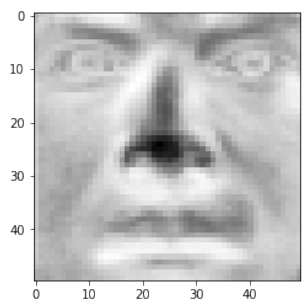
Sample test.



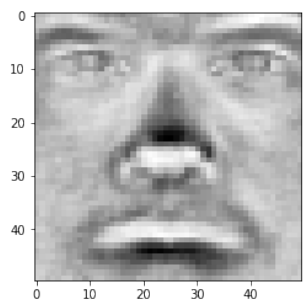
Sample train.



Sample test.

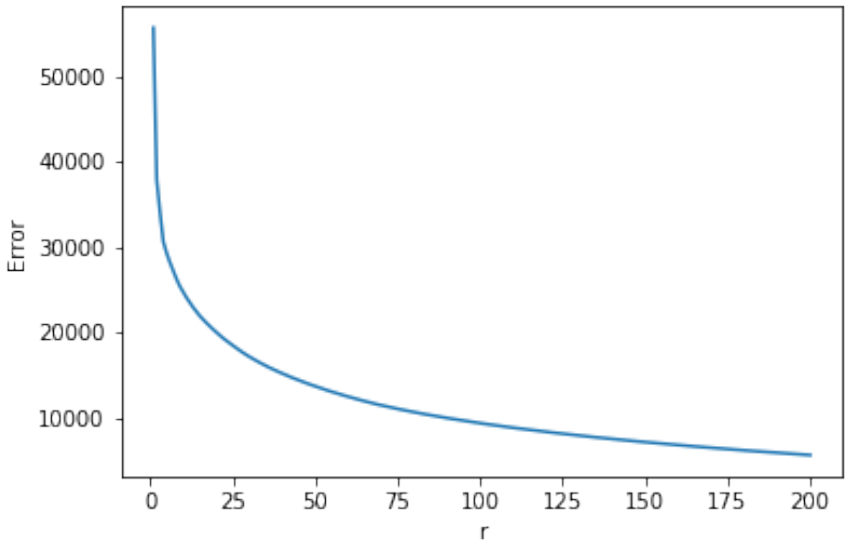


Sample train.



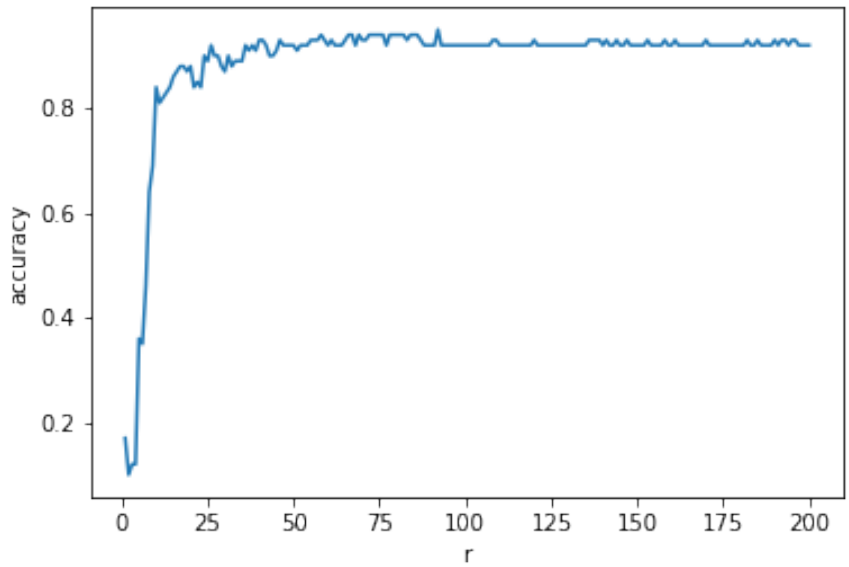
Sample test.

(f) Low-rank approximation error.



Rank r approximation error.

(h) Face recognition. Classification accuracy on test set with r=10 is 0.84.



Classification with Rank r approximated training data.

2. What's cooking?

(b) Training data.
Number of dishes: 39774
Number of unique ingredients: 6714
Number of cuisines: 20

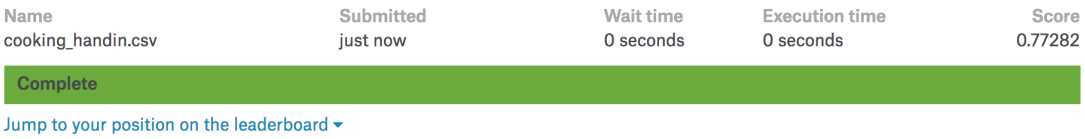
(d) Three-fold cross validation.
Gaussian prior: Average accuracy is 0.379846130638

Bernoulli prior: Average accuracy is 0.683536983959

(e) Bernoulli prior gave higher accuracy. This makes sense because we consider the presence or absence of an ingredient in a dish as features and encoded our features as binary values. It makes more sense to use Bernoulli distribution, which models discrete random variables taking only binary values. Gaussian distribution on the other hand is continuous over all real numbers. Our features are not normally distributed.

(e) Logistic Regression Average Accuracy: 0.775556896465

(f) Kaggle submission.



Kaggle submission.