

### ASSIGNEMENT-03

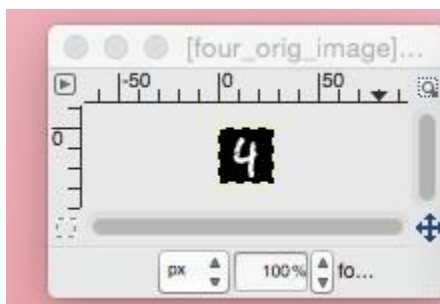
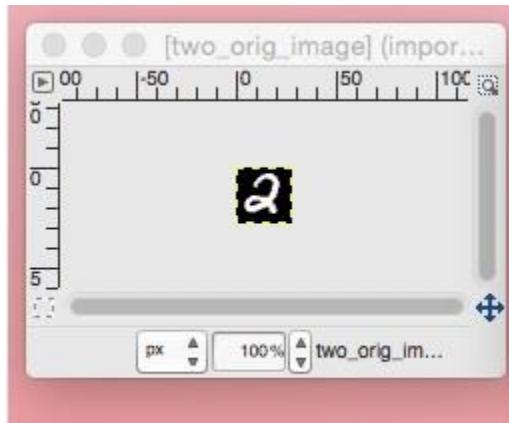
#### OBJECTIVE:

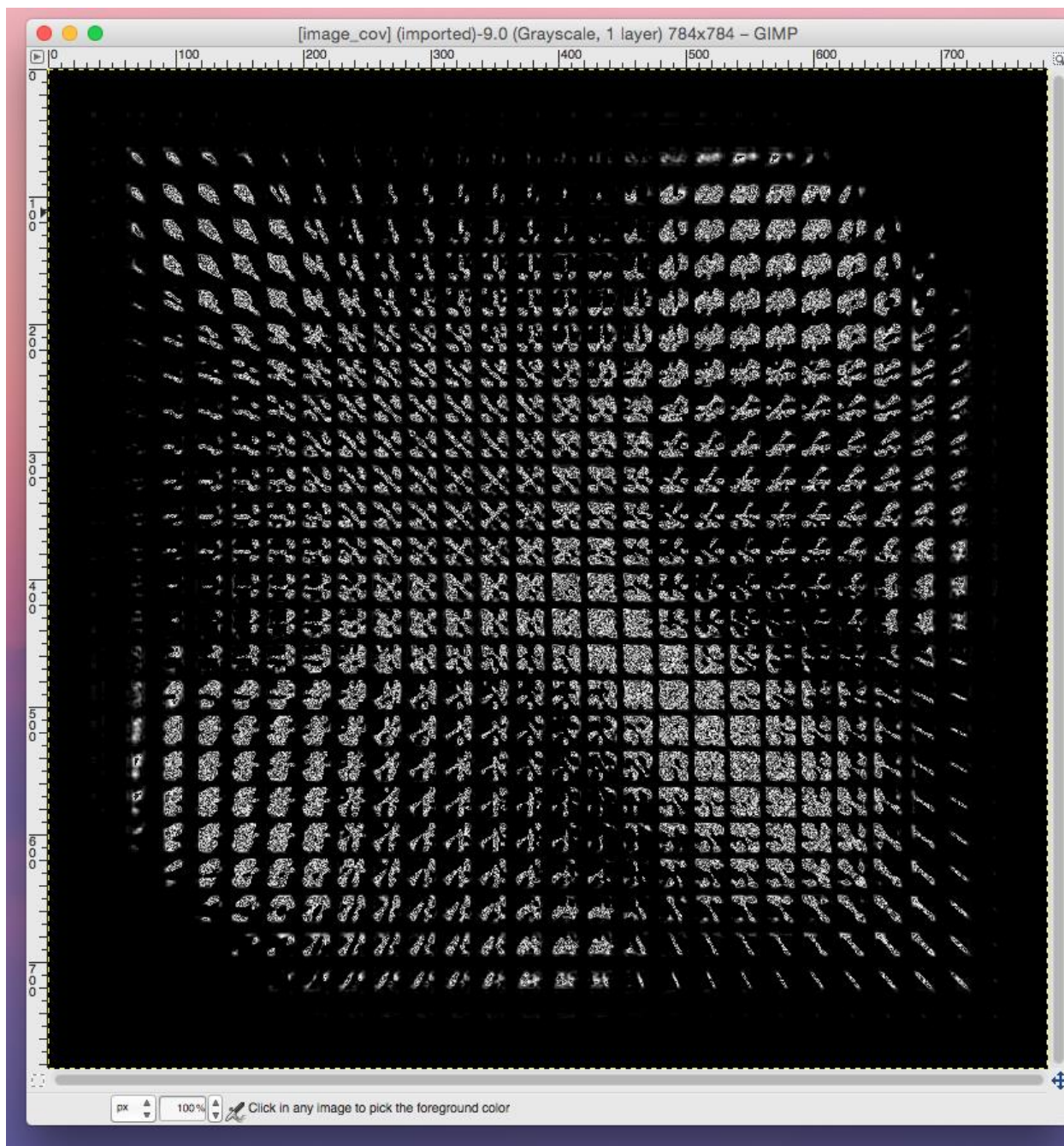
Perform following items.

1. Import handwritten digit data from [yann.lecun.com/exdb/mnist](http://yann.lecun.com/exdb/mnist)\*.
2. Write a program to convert any given 784 element feature vector to an image. See next slide.
3. Apply Principal Component Analysis to represent any two selected digits (*please do not choose digits 5 and 6*) in a reduced dimensional space e.g. 2D.
4. Visually verify that the reduction does not significantly compromise the identity of the samples
5. Build a simple OCR system to distinguish between the two digits. Build a histogram classifier and a Bayesian classifier (both  $\geq 2D$ ) assuming that the feature vectors are distributed normally in the reduced dimensional space.
6. Evaluate the performance of your classifiers using meaningful metrics.
7. Exhibit a few examples of correct and incorrect classification.

## TEST RESULTS

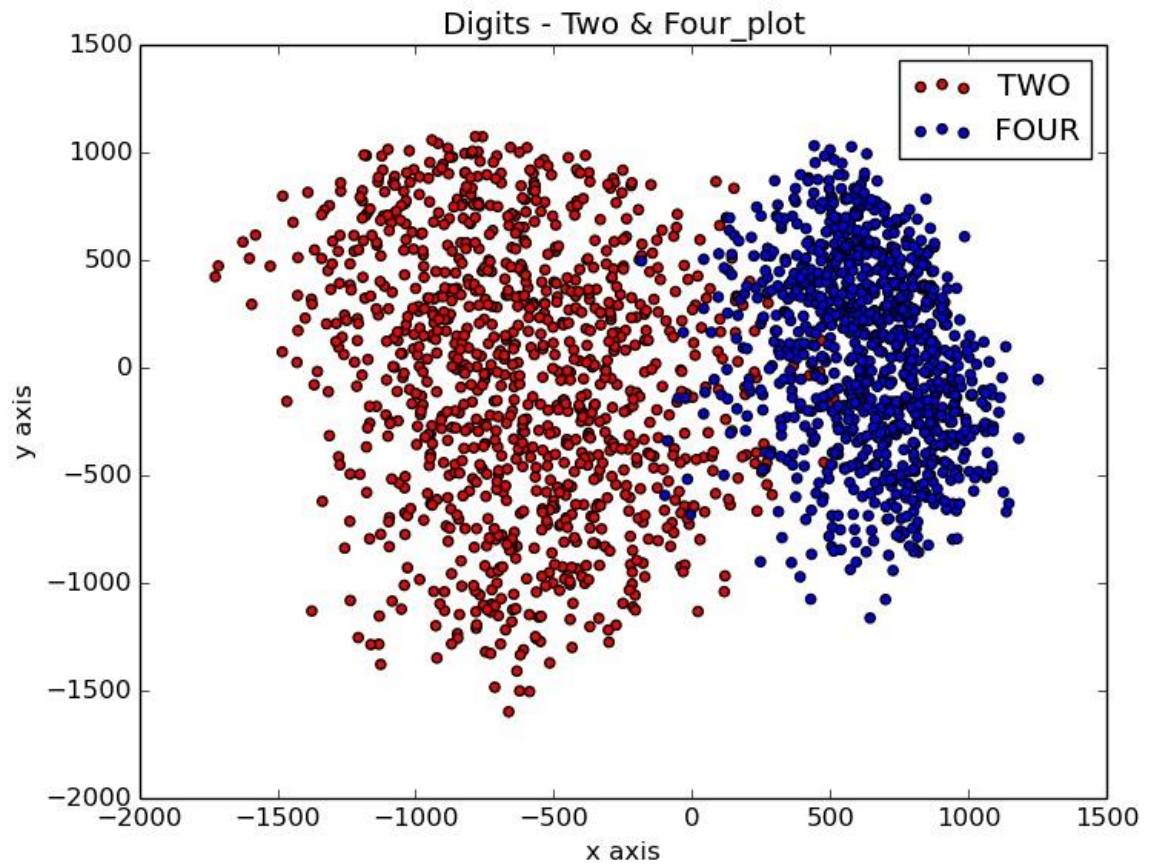
1. Import handwritten digit data from [yann.lecun.com/exdb/mnist\\*](http://yann.lecun.com/exdb/mnist).
  - Downloaded the files and read the data in Python.
  - READ only the data with 10K set, and constructed matching labels with 2 and 4 for my testing.
2. Program to write any 784 element to image.
  - Please find the sample images printed for numbers 2 & 4.
  - Also please find the co-variance matrix printed





3. Apply Principal Component Analysis to represent any two selected digits

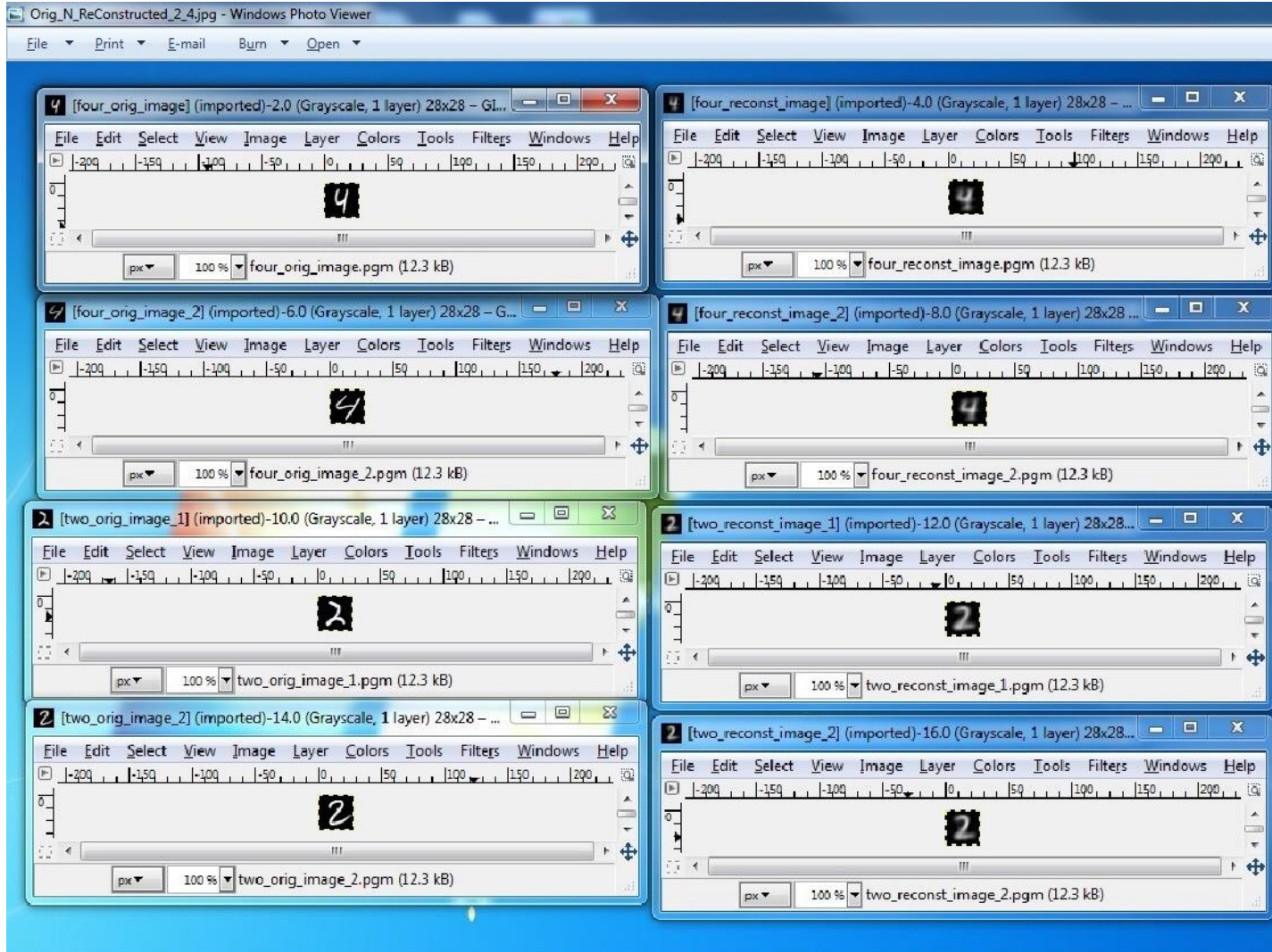
- Choose two digits as 2 and 4
- Please find the plot as per the PCA results.





4. Visually verify that the reduction does not significantly compromise the identity of the samples

Please find the digits TWO and FOUR before and after reduction in dimensions.



5. Build a simple OCR system to distinguish between the two digits. Build a histogram classifier and a Bayesian classifier (both  $\geq 2D$ ) assuming that the feature vectors are distributed normally in the reduced dimensional space.

- Build Bayesian classifier and
- Histogram

6. Evaluate the performance of your classifiers using meaningful metrics.

<b>ACCURACY</b>	<i>Bayesian</i>	0.9682224429
	Histogram	0.962717058223
<b>SENSITIVITY</b>	<i>Bayesian</i>	0.969450101833
	Histogram	0.988505747126
<b>SPECIFICITY</b>	<i>Bayesian</i>	0.967054263566
	Histogram	0.938061938062
<b>POSITIVE PREDICTIVE VALUE</b>	<i>Bayesian</i>	0.965517241379
	Histogram	0.938492063492

- Test for 100% match for digit-2

-

TEST DIGIT	2
No Of Test input	1032
Expected result	100 % MATCH

TEST RESULTS:

CLASSIFIER	RESULTS	
BAYES	MATCHED	998
	NOT MATCHED	34
	INDETERMINATE	0
HISTOGRAM	MATCHED	939
	NOT MATCHED	62
	INDETERMINATE	31

➤ Test for 100% match for digit-4

TEST DIGIT	4
No Of Test input	982
Expected result	100 % MATCH

TEST RESULTS:

CLASSIFIER	RESULTS	
BAYES	MATCHED	952
	NOT MATCHED	30
	INDETERMINATE	0
HISTOGRAM	MATCHED	946
	NOT MATCHED	30
	INDETERMINATE	0

➤ Test for 0% match for digit-1

TEST DIGIT	1
No Of Test input	1135
Expected result	0 % MATCH (No MATCH)

TEST RESULTS:

CLASSIFIER	RESULTS	
BAYES	MATCHED	1019
	NOT MATCHED	116
	INDETERMINATE	0
HISTOGRAM	MATCHED	415
	NOT MATCHED	153
	INDETERMINATE	567

➤ Test for 0% match for digit-6

TEST DIGIT	6
No Of Test input	1032
Expected result	0 % MATCH

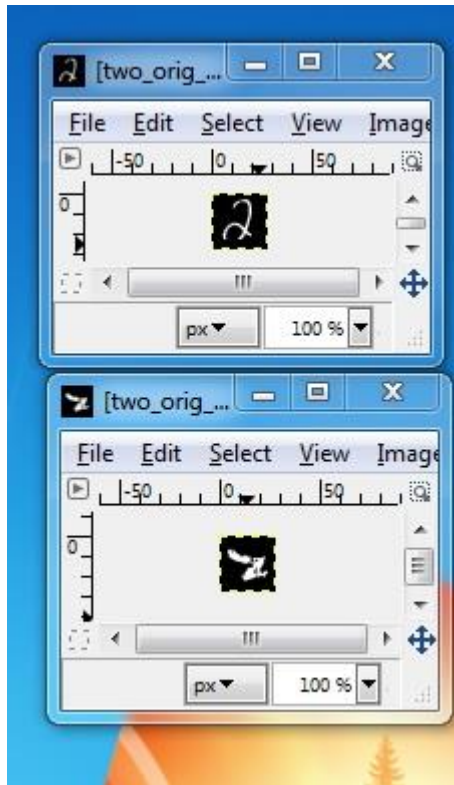
TEST RESULTS:

CLASSIFIER	RESULTS	
BAYES	MATCHED	1032
	NOT MATCHED	20
	INDETERMINATE	0
HISTOGRAM	MATCHED	946
	NOT MATCHED	11
	INDETERMINATE	25

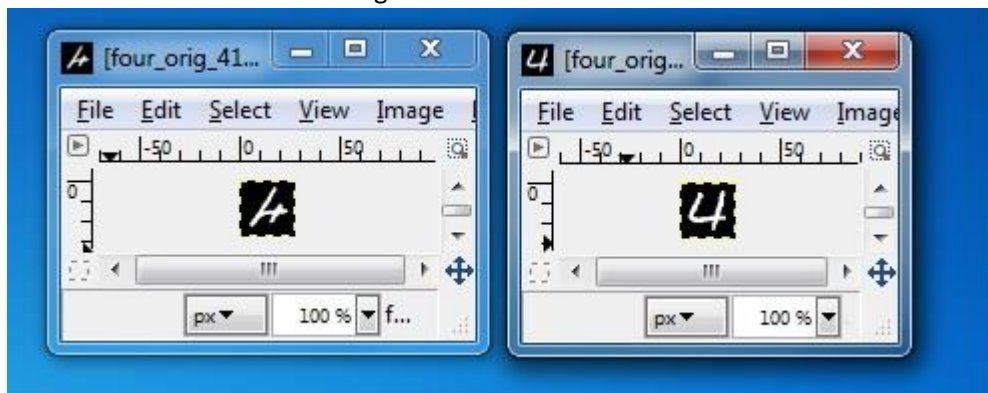


7. Exhibit a few examples of correct and incorrect classification.

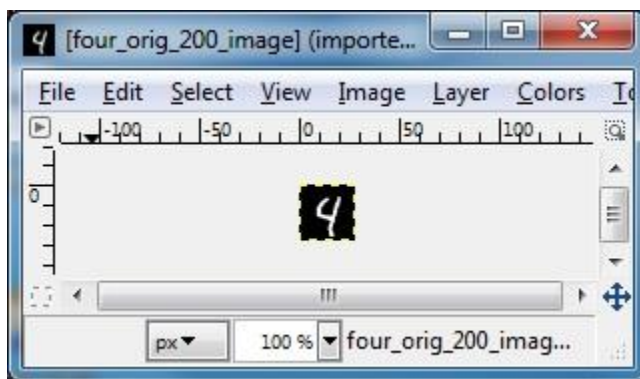
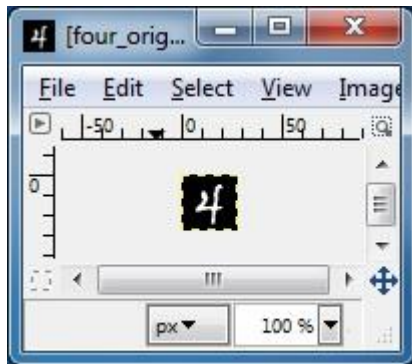
- INCORRECT classification for digit 2



- INCORRECT classification for digit 4



- Correct classification for digit 4



- Correct classification for digit 2

