

Image Analysis

Optimization and Data Analytics mini-project

Morten Lyng Rosenquist
Faculty of Technical Sciences
Aarhus University
Aarhus, Denmark
201706031

Write
ab-
stract

Abstract—

Index Terms— component, formatting, style, styling, insert

Insert
key-
words

I. INTRODUCTION

II. METHODS

III. DATASETS

The datasets that are analyzed in this paper are MNIST and ORL.

A. ORL

The ORL data set contains 400 vectorized images. The images are of size 40x30 and depict a persons face in an upright position in a frontal view[?]. There is 10 images of 40 different persons. Examples of the images can be seen on Figure 1.

Write
meth-
ods

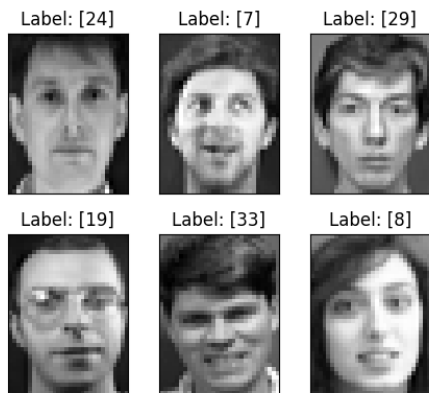


Fig. 1. ORL example images

B. MNIST

The MNIST dataset is larger with 70,000 vectorized images. The images depict handwritten digits and are of size 28x28. Differing from the ORL data set, the set is already split in training and test data. Leading to 60,000 images for training and 10,000 for testing. Examples from the MNIST data set can be seen on Figure 2.

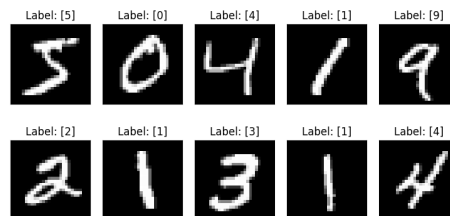


Fig. 2. MNIST example images

IV. RESULTS

The performance of each classifier for both datasets will in this section be covered.

A. PCA impact on data

Classification is performed both on the original data and on the data reduced to two dimensions using PCA. Illustrations of the data lost by reducing the dimensions can be seen on Figure 3 for ORL and Figure 4 for MNIST. These figures are the same images illustrated on Figure 1 and Figure 2, but where PCA has been applied and thereafter retransformed to their original dimensionality.

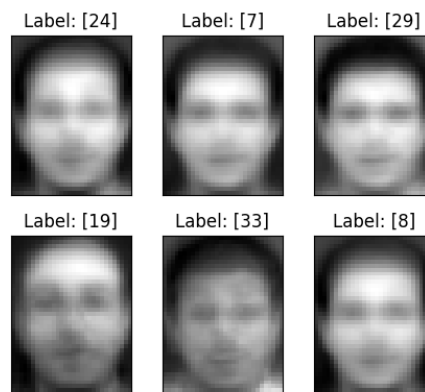


Fig. 3. ORL images reconstructed after PCA

Having the dimensions reduced to two, makes it easy to visualize the data. On Figure 5 can a scatter plot for the MNIST test data be seen.

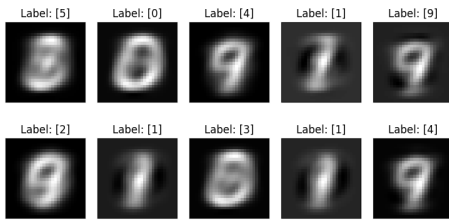


Fig. 4. MNIST images reconstructed after PCA

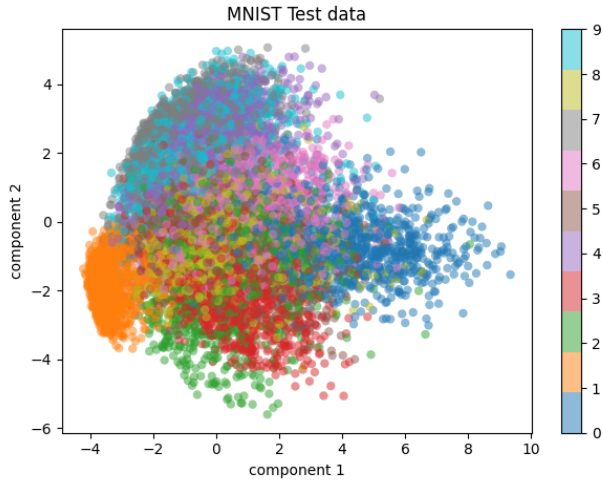


Fig. 5. MNIST test PCA images scatter plot

As for the ORL data set there is a 40 classes, which is hard to illustrate in a single scatter plot. Therefore Figure

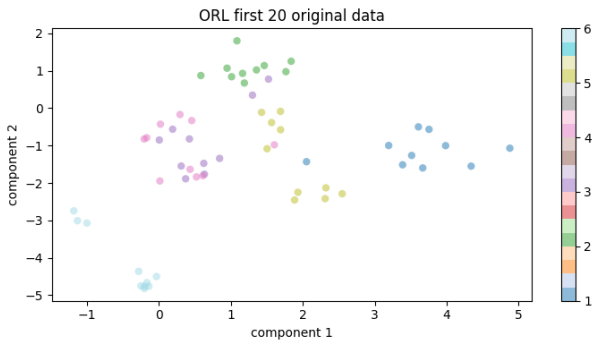


Fig. 6. MNIST test PCA images scatter plot

B. Classification

Each classifier, for both the original data and the PCA version of the data, have their hyperparameters tuned if relevant. Afterwards the classification is performed. The accuracy of class predictions of the test data can be seen on Table I. Measurements of the time spent both training and testing can be seen for certain classifiers.

Dataset	Classifier	Accuracy (Raw)	Accuracy (2d)	Time
MNIST	Nearest Class Centroid	1%	2%	2 m
	Nearest 2 Sub-Class Centroid	1%		
	Nearest 3 Sub-Class Centroid	1%		
	Nearest 5 Sub-Class Centroid	15		
	Nearest Neighbor			
	Perceptron with Backpropagation			
	Perceptron with MSE			
ORL	Nearest Class Centroid			
	Nearest 2 Sub-Class Centroid			
	Nearest 3 Sub-Class Centroid			
	Nearest 5 Sub-Class Centroid			
	Nearest Neighbor			
	Perceptron with Backpropagation			
	Perceptron with MSE			

TABLE I
PERFORMANCE OF EACH CLASSIFIER FOR BOTH DATASETS

C. Visualization

V. CONCLUSION

Write
con-
clu-
sion

NOTES

<input type="checkbox"/>	Write abstract	1
<input type="checkbox"/>	Insert keywords	1
<input type="checkbox"/>	Write introduction	1
<input type="checkbox"/>	Write methods	1
<input type="checkbox"/>	Write conclusion	2