

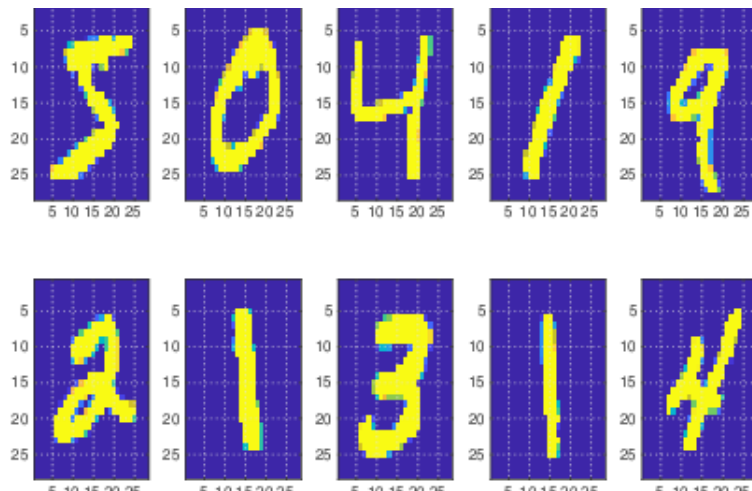
Machine learning algorithms for inverse problems

Course Project

Larisa Beilina, larisa@chalmers.se

INSTRUCTIONS

- You can work in groups by 2 persons.
- Sent final report for every computer assignment with description of your work together with Matlab or C++/PETSc programs to my e-mail before deadline. Report should have description of used techniques, tables and figures confirming your investigations. Analysis of obtained results is necessary to present in section “Numerical examples” and summarized results - in section “Conclusions”. You can download latex or pdf-template for report from the course homepage.



COURSE PROJECT

PRINCIPAL COMPONENT ANALYSIS FOR IMAGE RECOGNITION

For Master's students: this exercise can be viewed as background for the Master's project "Applications of Principal Component Analysis in computer vision"

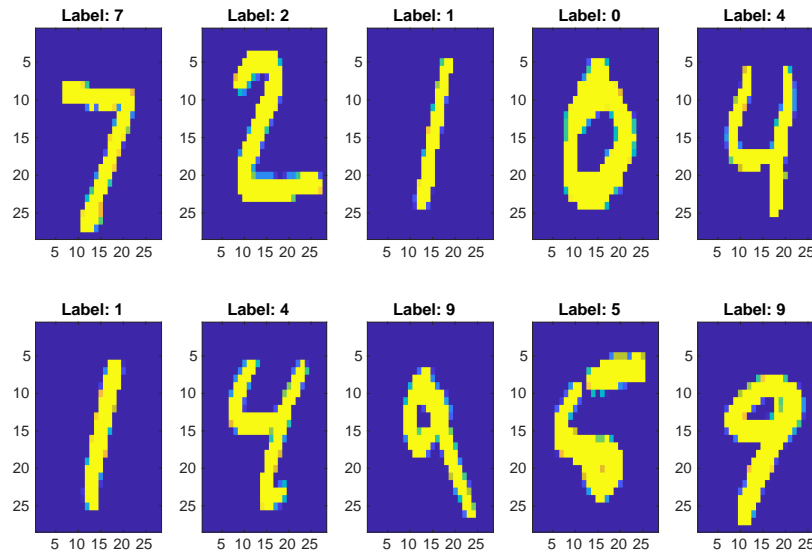


Figure 0.1: Images from MNIST dataset `mnist_test_10.csv` visualised via the program `loadmnist_matlab.m`

- Principal component analysis (PCA) is a machine learning technique which is widely used for data compression in image processing (data visualization) or in the determination of object orientation.
- PCA problem is closely related to the numerical linear algebra (NLA) problem of finding eigenvalues and eigenvectors for the covariance matrix.
- Further reading for AI algorithms: [1, 2, 3].

Project assignments

- 1. Use PCA to find patterns (recognize handwritten numbers) in MNIST Dataset of Handwritten Digits which can be downloaded from the link

<http://makeyourownneuralnetwork.blogspot.com/2015/03/the-mnist-dataset-of-handwritten-digits.html>

or from the course homepage.

- 2. USE PCA to classify skin images from the ISIC project, see link
<https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main>

Hints:

- Study lecture about PCA and the Matlab program which performs PCA for 2 datasets

```
ExamplePCA.m
```

on the course homepage.

- Use Matlab programs

```
loadmnist_matlab.m
import_mnist.m
```

on the course homepage to download MNIST Datasets

```
mnist_test_10.csv
mnist_train.csv
```

- Perform PCA analysis for the following problem: given an image from the dataset

```
mnist_test_10.csv
```

check if there exists the same image in the dataset

```
mnist_train.csv
```

- Use PCA to classify skin images from the ISIC project, see link

<https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main>

Choose test dataset from ISIC database and compare with other images of this dataset in order to determine to which one class belongs image from train dataset. During classification assume that you don't know true class of images from train dataset. Compare then classified images with true ones.

- Compute missclassification rate E using the formula (see [3], p. 211-214):

$$E = \frac{\sum_{i=1}^K N_{F,i}}{\sum_{i=1}^K (N_{T,i} + N_{F,i})}, \quad (0.1)$$

where K is the number of classes, $N_{T,i}$ is the number of images of the class i which are classified correctly, $N_{F,i}$ is the number of images of the class i which are classified wrong. Precision for class i can be computed as

$$P(i) = \frac{N_{T,i}}{N_{T,i} + N_{F,j}}. \quad (0.2)$$

REFERENCES

- [1] Christopher M. Bishop, *Pattern recognition and machine learning*, Springer, 2009.
- [2] Ian Goodfellow, Yoshua Bengio and Aaron Courville, *Deep Learning*, MIT Press, 2016,
<http://www.deeplearningbook.org>
- [3] Miroslav Kurbat, *An Introduction to Machine Learning*, Springer, 2017.