

[Overview](#)[Data](#)[Kernels](#)[Discussion](#)[Leaderboard](#)[Rules](#)

Overview

Description

Description

Evaluation

This is an example classification problem based on the historic Olivetti Dataset put together by AT&T Laboratories around 1994. The data consists of 400 images of 40 distinct people captured by 64x64 grey-scale images [6]. While more modern and large-scale facial datasets are currently available, the Olivetti Dataset provides a small dataset tractable for doing a few exercises and also provides some historical context [1-4] for understanding modern state-of-the-art approaches [5,6].

In this exercise, we show how to load the image data and perform pre-processing using Principle Component Analysis (PCA) to extract features (eigen-faces [1]) for a reduced description of the facial images. You will then perform classification using these eigen-face features to get experience with Support Vector Machines (SVMs). Compare how different choices of dimension reduction (number of features retained) and other hyper-parameters influence classification. For further discussions of both historic and modern approaches to facial recognition see the references [1-6].

We remark that one also could use modern techniques to perform inference more directly on the images, such as Convolutional Neural Networks (CNNs), which we shall discuss more in later lectures. While you are also welcome as an exercise to develop a CNN to perform classification as a point of comparison, one challenge here is the small size of the Olivetti Dataset which might be insufficient for learning adequately from scratch a good CNN, but you could try transfer learning from a pre-trained CNN.

The prediction to submit is the class of the person identified in the given test images. Use the class indices labelled between 0 to 39 for the 40 distinct people in the dataset.

References

[1] *Eiegenfaces for Recognition*, Turk, N. and Pentland, A.P., J. Cognitive Neuroscience, (1991).

[2] *The Model Method in Facial Recognition*, Bledsoe, W. W., Technical Report PRI 15, Panoramic Research, Inc., Palo Alto, California, (1964).

[3] *Semiautomatic Facial Recognition*, Bledsoe, W. W., Technical Report SRI Project 6693, Stanford Research Institute, Menlo Park, California, (1968).

[4] *Computer Recognition of Human Faces*, Kanade, T., Interdisciplinary Systems Research, 47, January, (1977).

[5] *Comparison of human and computer performance across face recognition experiments*, Phillips, P., O'Toole, A., Image and Vision Computing 32, (2014).

[6] *Deep learning*, LeCun, Y., Bengio, Y., and Hinton, G., Nature, 521, May, (2015).

[7] Images from AT&T Laboratories Cambridge.

<https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

Author of Competition

This competition was setup for a data science course by Paul J. Atzberger, <http://atzberger.org/>



[Overview](#)

[Data](#)

[Kernels](#)

[Discussion](#)

[Leaderboard](#)

[Rules](#)

Overview

Description

Evaluation

[Evaluation](#)

For this classification task your methods will be scored by the number of correct labels you predicted as determined from the test data set.

Launch
5 days ago

Close
2 months



[Overview](#)[Data](#)[Kernels](#)[Discussion](#)[Leaderboard](#)[Rules](#)

Data Description

This is the Olivetti Dataset consists of a collection of faces for recognition tasks. The images are grey-scale 64x64 images. They are arranged in a .npy (numpy archive format) as nx64x64 tensor. To get image I one uses numpy notation `image = data[I]`.

File descriptions

- `train_data.npy` - the training data for the face images
- `train_labels.npy` - the training labels for the face images
- `test_data.npy` - the test subset of face images

The data set consists of 64x64 images for which you should make a prediction of the labels. Submit your labels for the `test_data.npy` data set in the format of `submission_sample.csv` for kaggle scoring.

Data (3 MB)

[API](#)[kaggle competitions download -c atz-facial-rec-e...](#)[Download All](#)

Data Sources

submission_samp...	80 x 2
--------------------	--------

[olivetti_faces.npy](#)[olivetti_faces_target.npy](#)[test_data.npy](#)[train_data.npy](#)[train_labels.npy](#)

About this file

No description yet

Columns

id

label