COM6018 Data Science with Python

Lab 7: Classification with Scikit-Learn

Jon Barker

In this lab

Using Scikit-Learn to compare different classifiers.

- Preparing a face recognition task
- Perform feature pre-processing and dimensionality reduction
- Building k-NN, SVM, Random Forest and MLP classifiers
- Using GridSearch to perform parameter optimising
- Using pipelines to streamline the evaluation process
- Evaluating using confusion matrices and per-class precision, recall and F1 scores.

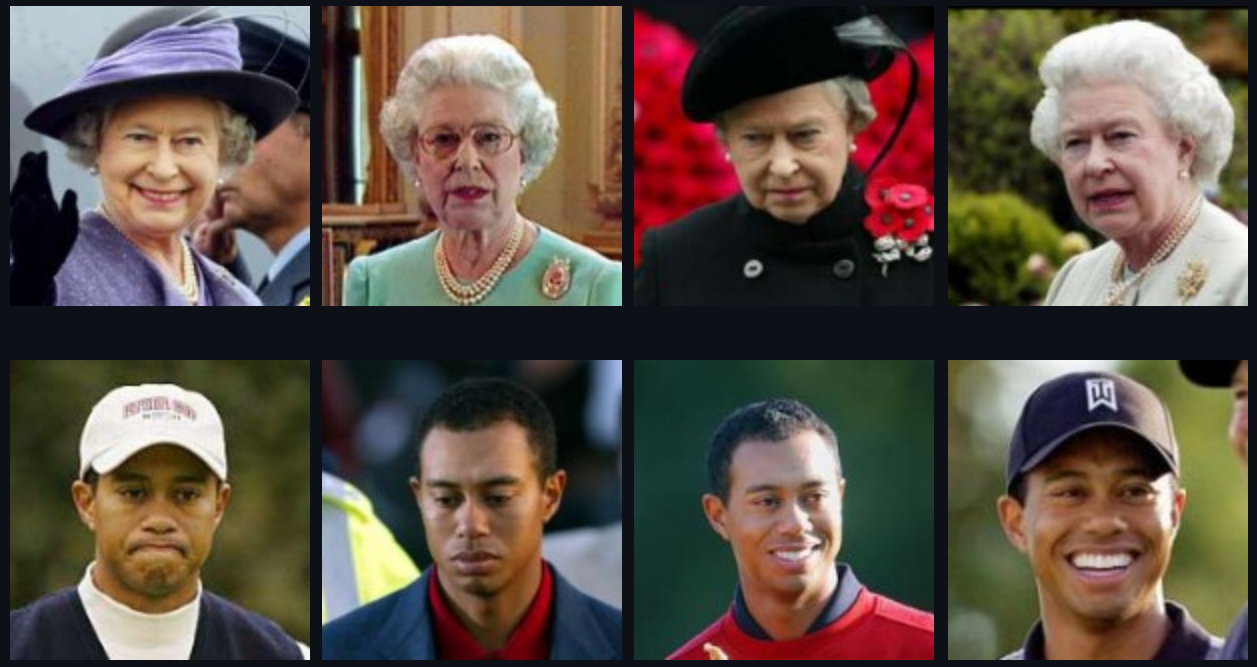
The Task

We will use Scikit-Learn to recognise famous people from photographs.

- We will use Scikit Learn's builtin data set, 'Labeled Faces in the Wild'
- Details here, http://vis-www.cs.umass.edu/lfw/
- It contains 13,233 images of 5749 famous people.
- Designed for a face verification task, but we will use a subset of it for a classification task.

You will be using the same data in Assignment 2, so this is a good opportunity to get familiar with it.

Example Data



Examples from the 'Labeled Faces in the Wild' dataset.

Example Data

Examples from the 'Labeled Faces in the Wild' dataset.

The Data

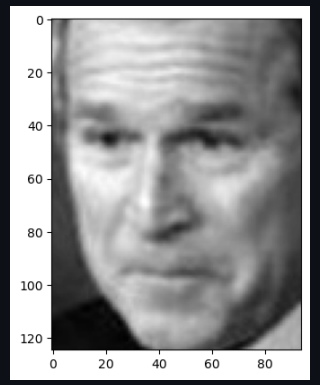
dict_keys(['data', 'images', 'target', 'target_names', 'DESCR'])

images - 3D numpy array storing orignal colour images (N x height x width)
data - 2D numpy array storing pre-processed images (N x n_pixels)
target - a list of the N labels (stored as integers)
target_names - the names associated with each class

Preprocessing

The data has been pre-processed

- images cropped to the face.
- transformed to grey-scale.
- resized to 125x94 pixels.



Obtaining the Jupyter Notebook

If you have cloned and pulled the module's GitHub repository then you should see,

```
materials/labs/
├── 070_classification_with_scikit_learn.ipynb
|-- ... etc
```

The lab is `070_classification_with_scikit_learn.ipynb`. It does not require any additional data files. (The dataset is built into Scikit-Learn.)

Or you can download the notebook and data via links on Blackboard.

Getting Help

- If you are stuck just raise a hand to ask for help.
- Feel free to discuss the lab with your neighbours.
- Re-read the Scikit-Learn tutorial notes
    - In the Git repo at `materials/tutorials/070_Classification_with_Scikit_Learn.ipynb`
    - or online at https://uos-com-6018.github.io/COM6018
- Use the Scikit-Learn API documentation for reference. https://scikit-learn.org/stable/modules/classes.html