

```
In [1]: import numpy,sklearn,sklearn.datasets,utils
        %matplotlib inline
```

Principal Component Analysis

In this exercise, we will experiment with two different techniques to compute the PCA components of a dataset:

- **Singular Value Decomposition (SVD)**
- **Power Iteration:** A technique that iteratively optimizes the PCA objective.

We consider a random subset of the Labeled Faces in the Wild (LFW) dataset, readily accessible from sklearn, and we apply some basic preprocessing to discount strong variations of luminosity and contrast.

```
In [2]: D = sklearn.datasets.fetch_lfw_people(resize=0.5)['images']
        D = D[numpy.random.mtrand.RandomState(1).permutation(len(D))[:2000]]*1.0
        D = D - D.mean(axis=(1,2),keepdims=True)
        D = D / D.std(axis=(1,2),keepdims=True)
        print(D.shape)
```

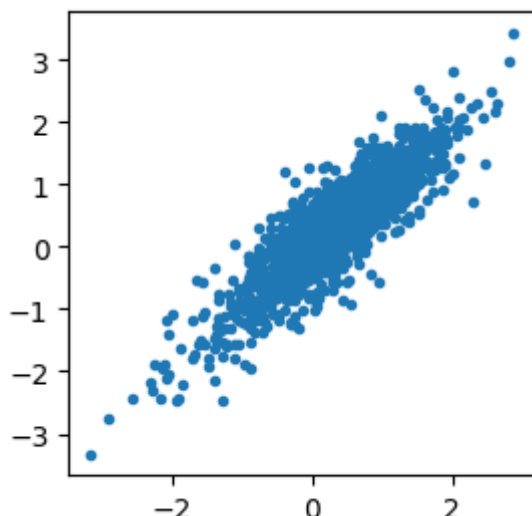
(2000, 62, 47)

Two functions are provided for your convenience and are available in `utils.py` that is included in the zip archive. The functions are the following:

- **utils.scatterplot** produces a scatter plot from a two-dimensional data set.
- **utils.render** takes an array of data points or objects of similar shape, and renders them in the IPython notebook.

Some demo code that makes use of these functions is given below.

```
In [3]: utils.scatterplot(D[:,32,20],D[:,32,21]) # Plot relation between adjacent pixels
        utils.render(D[:30],15,2,vmax=5)         # Display first 10 examples in the data
```





PCA with Singular Value Decomposition (15 P)

Principal components can be found computing a singular value decomposition. Specifically, we assume a matrix \bar{X} whose columns contain the data points represented as vectors, and where the data points have been centered (i.e. we have subtracted to each of them the mean of the dataset). The matrix \bar{X} is of size $d \times N$ where d is the number of input features and N is the number of data points. This matrix, more specifically, the rescaled matrix $Z = \frac{1}{\sqrt{N}} \bar{X}$ is then decomposed using singular value decomposition:

$$U \Lambda V = Z$$

The k principal components can then be found in the first k columns of the matrix U .

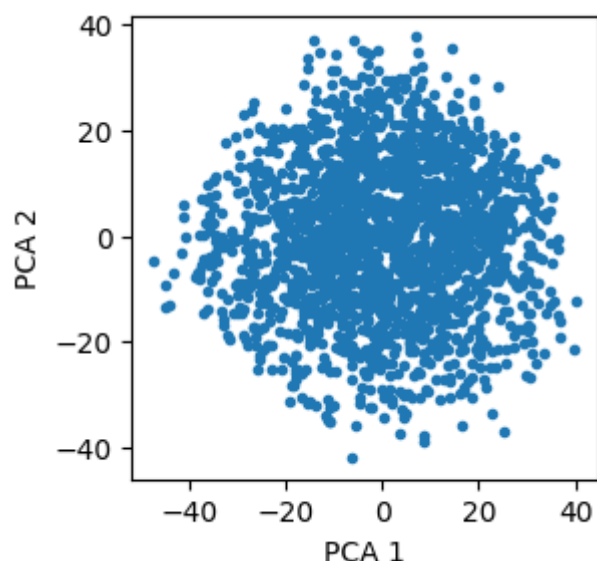
Tasks:

- Compute the principal components of the data using the function `numpy.linalg.svd`.
- Measure the computational time required to find the principal components. Use the function `time.time()` for that purpose. Do *not* include in your estimate the computation overhead caused by loading the data, plotting and rendering.
- Plot the projection of the dataset on the first two principal components using the function `utils.scatterplot`.
- Visualize the 60 leading principal components using the function `utils.render`.

Note that if the algorithm runs for more than 3 minutes, there may be some error in your implementation.

```
In [4]: ### REPLACE BY YOUR CODE
import solutions; solutions.pca_svd(D)
###
```

Time: 5.428 seconds





When looking at the scatter plot, we observe that much more variance is expressed in the first two principal components than in individual dimensions as it was plotted before. When looking at the principal components themselves which we render as images, we can see that the first principal components correspond to low-frequency filters that select for coarse features, and the following principal components capture progressively higher-frequency information and are also becoming more noisy.

PCA with Power Iteration (15 P)

The first PCA algorithm based on singular value decomposition is quite expensive to compute. Instead, the power iteration algorithm looks only for the first component and finds it using an iterative procedure. It starts with an initial weight vector $\mathbf{w} \in \mathbb{R}^d$, and repeatedly applies the update rule

$$\mathbf{w} \leftarrow S\mathbf{w} / \|S\mathbf{w}\|.$$

where S is the covariance matrix defined as $S = \frac{1}{N} \bar{X} \bar{X}^\top$. Like for standard PCA, the objective that iterative PCA optimizes is $J(\mathbf{w}) = \mathbf{w}^\top S \mathbf{w}$ subject to the unit norm constraint for \mathbf{w} . We can therefore keep track of the progress of the algorithm after each iteration.

Tasks:

- **Implement the power iteration algorithm. Use as a stopping criterion the value of $J(\mathbf{w})$ between two iterations increasing by less than 0.01.**
- **Print the value of the objective function $J(\mathbf{w})$ at each iteration.**
- **Measure the time taken to find the principal component.**
- **Visualize the the eigenvector \mathbf{w} obtained after convergence using the function `utils.render`.**

Note that if the algorithm runs for more than 1 minute, there may be some error in your implementation.

```
In [5]: ### REPLACE BY YOUR CODE
import solutions; solutions.pca_powit(D)
###
```

```
iteration 0    J(w) =      0.667
iteration 1    J(w) =    122.218
iteration 2    J(w) =    201.633
iteration 3    J(w) =    215.603
iteration 4    J(w) =    224.944
iteration 5    J(w) =    234.307
iteration 6    J(w) =    243.160
iteration 7    J(w) =    250.604
iteration 8    J(w) =    256.206
iteration 9    J(w) =    260.069
iteration 10   J(w) =    262.571
iteration 11   J(w) =    264.128
iteration 12   J(w) =    265.072
iteration 13   J(w) =    265.635
iteration 14   J(w) =    265.969
iteration 15   J(w) =    266.165
iteration 16   J(w) =    266.280
iteration 17   J(w) =    266.347
iteration 18   J(w) =    266.386
iteration 19   J(w) =    266.409
iteration 20   J(w) =    266.423
iteration 21   J(w) =    266.431
stopping criterion satisfied
Time: 0.675 seconds
```



We observe that the computation time has decreased significantly. The difference of performance becomes larger as the number of dimensions and data points increases. We can observe that the principal component is the same (sometimes up to a sign flip) as the one obtained by the SVD algorithm.