

# PCA-Algorithms-Excercise

February 17, 2021

## 1 Objective: To develop a mathematical understanding of PCA

## 2 Test data

```
[1]: import numpy as np
from scipy import linalg as LA
X = np.array([
    [0.387, 4878, 5.42],
    [0.723, 12104, 5.25],
    [1, 12756, 5.52],
    [1.524, 6787, 3.94],
])
X = X - np.mean(X, axis=0)
```

## 3 Non-linear Iterative Partial Least-Squares (NIPALS) algorithm

Steps to compute PCA using NIPALS algorithm

- Step 1: Initialize an arbitrary column vector  $\mathbf{t}_a$  either randomly or by just copying any column of  $X$ .
- Step 2: Take every column of  $X$ ,  $X_k$  and regress it onto the  $\mathbf{t}_a$  vector and store the regression coefficients as  $p_{ka}$ . (Note: This simply means performing an ordinary least squares regression ( $y = mx$ ) with  $x = t_a$  and  $y = X_k$  with  $m = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{y}$ ). In the current notation we get

$$p_{ka} = \frac{\mathbf{t}_a^T X_k}{\mathbf{t}_a^T \mathbf{t}_a}$$

Repeat it for each of the columns of  $X$  to get the entire vector  $\mathbf{p}_k$ . This is shown in the illustration above where each column from  $X$  is regressed, one at a time, on  $\mathbf{t}_a$ , to calculate the loading entry,  $p_{ka}$

In practice we don't do this one column at a time; we can regress all columns in  $X$  in go:

$$\mathbf{p}_a^T = \frac{1}{\mathbf{t}_a^T \mathbf{t}_a} \cdot \mathbf{t}_a^T X_a$$

where  $\mathbf{t}_a$  is an  $N \times 1$  column vector, and  $\mathbf{X}_a$  us an  $N \times K$  matrix. \* The loading vector  $\mathbf{p}_a^T$  won't have unit length (magnitude) yet. So we simply rescale it to have magnitude of 1.0:

$$\mathbf{p}_a^T = \frac{\mathbf{p}_a^T}{\sqrt{\mathbf{p}_a^T \mathbf{p}_a}}$$

\* Step 4: Regress every row in  $X$  onto this normalized loadings vector. As illustrated below, in our linear regression the rows in  $X$  are our y-variable each time, while the loadings vector is our x-variable. The regression coefficient becomes the score value for that  $^{th}$  row:

$$p_{i,a} = \frac{\mathbf{x}_i^T \mathbf{p}_a}{\mathbf{p}_a^T \mathbf{p}_a}$$

where  $\mathbf{x}_i^T$  is a  $K \times 1$  column vector. We can combine these  $N$  separate least-squares models and calculate them in one go to get the entire vector,

$$\mathbf{t}_a^T = \frac{1}{\mathbf{p}_a^T \mathbf{p}_a} \mathbf{X} \mathbf{p}_a^T$$

where  $p_a$  is a  $K \times 1$  column vector. \* Step 5: Continue looping over steps 2,3,4 until the change in vector  $t_a$  is below a chosen tolerance \* Step 6: On convergence, the score vector and the loading vectors,  $\mathbf{t}_a$  and  $\mathbf{p}_a$  are stored as the  $a^{th}$  column in matrix  $\mathbf{T}$  and  $\mathbf{P}$ . We then deflate the  $\mathbf{X}$  matrix. This crucial step removes the variability captured in this component ( $t_a$  and  $p_a$ ) from  $\mathbf{X}$ :

$$E_a = X_a - t_a p_a^T$$

$$X_{a+1} = E_a$$

For the first component,  $X_a$  is just the preprocessed raw data. So we can see that the second component is actually calculated on the residuals  $E_1$ , obtained after extracting the first component. This is called deflation, and nicely shows why each component is orthogonal to the others. Each subsequent component is only seeing variation remaining after removing all the others; there is no possibility that two components can explain the same type of variability. After deflation we go back to step 1 and repeat the entire process for the next component.

### 3.1 IMPLEMENTATION IN PYTHON

```
[2]: def PCA(X,no_components):
    tol = 0.0000001
    it=1000
    obsCount,varCount = X.shape
    Xa = X - np.mean(X, axis = 0)
    #Xh = X-np.tile(np.mean(X,axis=0).reshape(-1,1).T, obsCount).reshape(4,3)
    T = np.zeros((obsCount,no_components))
    P = np.zeros((varCount,no_components))
    pcvar = np.zeros((varCount,1))
    varTotal = np.sum(np.var(Xa,axis=0,ddof=1))
    currVar = varTotal
    nr=0
    for h in range(no_components):
```

```

th = Xa[:,0].reshape(-1,1)
ende = False
while ende != True:
    nr = nr + 1
    ph = np.dot(Xa.T,th)/np.dot(th.T,th)
    ph = ph /np.linalg.norm(ph)
    thnew = np.dot(Xa,ph)/np.dot(ph.T,ph)
    prec = np.dot((thnew-th).T,(thnew-th))
    th = thnew
    if prec <= (tol*tol):
        ende = True
    elif it <=nr:
        ende = True
        print("Iteration stops without convergence")
Ea = Xa - np.dot(th,ph.T)
Xa = Ea
T[:,h] = th.flatten()
P[:,h] = ph.flatten()
oldVar = currVar
currVar = np.sum(np.var(Xa,axis=0,ddof=1))
pcvar[h] = (oldVar - currVar) / varTotal
nr = 0
return T,P,pcvar

```

### 3.2 Advantages of the NIPALS algorithm

- The NIPALS algorithm computes one component at a time. The first component computed is equivalent to the t1 and p1 vectors that would have been found from an eigenvalue or singular value decomposition.
- The algorithm can handle missing data in X.
- The algorithm always converges, but the convergence can sometimes be slow.
- It is also known as the Power algorithm to calculate eigenvectors and eigenvalues.
- It works well for very large data sets.
- It is used by most software packages, especially those that handle missing data.
- Of interest: it is well known that Google used this algorithm for the early versions of their search engine, called PageRank148.

```

[3]: no_components=3
T,P,pcvar = PCA(X,no_components)
print("T (Scores)")
print(T)
print(" ")
print("P (Loadings)")
print(P)
print(np.sqrt(pcvar)/np.sum(np.sqrt(pcvar)))

```

T (Scores)

```

[[-4.25324997e+03 -8.41288672e-01  8.37859036e-03]
 [ 2.97275001e+03 -1.25977272e-01 -1.82476780e-01]
 [ 3.62475003e+03 -1.56843494e-01  1.65224286e-01]
 [-2.34425007e+03  1.12410944e+00  8.87390330e-03]]

```

P (Loadings)

```

[[ 1.21901390e-05  5.66460728e-01  8.24088735e-01]
 [ 9.99999997e-01  5.32639787e-05 -5.14047689e-05]
 [ 7.30130279e-05 -8.24088733e-01  5.66460726e-01]]
[9.99753412e-01]
[2.10083377e-04]
[3.65048880e-05]]

```

## 4 SVD

```

[4]: from numpy.linalg import svd
      U, S, PTrans = svd(X, full_matrices=False)
      Sigma = np.diag(S)
      T=np.dot(U,Sigma)
      P=PTrans.T

      print("T (Scores)")
      print(T)
      print(" ")
      print("P (Loadings)")
      print(P)
      print("Sigma (Variance)")
      print(S)

```

T (Scores)

```

[[-4.25324997e+03  8.41288672e-01 -8.37858943e-03]
 [ 2.97275001e+03  1.25977271e-01  1.82476780e-01]
 [ 3.62475003e+03  1.56843494e-01 -1.65224286e-01]
 [-2.34425007e+03 -1.12410944e+00 -8.87390454e-03]]

```

P (Loadings)

```

[[ 1.21901390e-05 -5.66460727e-01 -8.24088736e-01]
 [ 9.99999997e-01 -5.32639789e-05  5.14047691e-05]
 [ 7.30130279e-05  8.24088734e-01 -5.66460725e-01]]

```

Sigma (Variance)

```

[6.74994067e+03  1.41840009e+00  2.46466604e-01]

```

## 5 SKLEARN PCA

```
[5]: from sklearn.decomposition import PCA
pca = PCA()
T=pca.fit_transform(X)
Prans=pca.components_ #eigen vectors.T
latent = pca.explained_variance_
explained = pca.explained_variance_ratio_
P=PTrans.T
S=pca.singular_values_
Sigma=np.diag(S)
print("T (Scores)")
print(T)
print(" ")
print("P (Loadings)")
print(P)
print("Sigma (Variance)")
print(S)
#print(pca.singular_values_/np.sqrt(3))
```

T (Scores)

```
[[ 4.25324997e+03 -8.41288672e-01 -8.37858943e-03]
 [-2.97275001e+03 -1.25977271e-01  1.82476780e-01]
 [-3.62475003e+03 -1.56843494e-01 -1.65224286e-01]
 [ 2.34425007e+03  1.12410944e+00 -8.87390454e-03]]
```

P (Loadings)

```
[[ 1.21901390e-05 -5.66460727e-01 -8.24088736e-01]
 [ 9.99999997e-01 -5.32639789e-05  5.14047691e-05]
 [ 7.30130279e-05  8.24088734e-01 -5.66460725e-01]]
```

Sigma (Variance)

```
[6.74994067e+03  1.41840009e+00  2.46466604e-01]
```

```
[6]: pca.explained_variance_ratio_
```

```
[6]: array([9.99999955e-01, 4.41567976e-08, 1.33326424e-09])
```

```
[7]: explained_variance_2 = (S ** 2) / 4
explained_variance_ratio_2 = (explained_variance_2 / explained_variance_2.sum())
print(explained_variance_ratio_2)
```

```
[9.99999955e-01  4.41567976e-08  1.33326424e-09]
```

## 6 Eigenvalue decomposition approach

Recall that the latent variable directions (the loading vectors) were oriented so that the variance of the scores in that direction were maximal. We can cast this as an optimization problem. For the

first component:

$$\max(\phi) = \mathbf{t}_1^T \mathbf{t}_1 = \mathbf{p}_1^T \mathbf{X}^T \mathbf{X} \mathbf{p}_1$$

such that

$$\mathbf{p}_1^T \mathbf{p}_1 = 1$$

This is equivalent to

$$\max(\phi) = \mathbf{p}_1^T \mathbf{X}^T \mathbf{X} \mathbf{p}_1 - \lambda(\mathbf{p}_1^T \mathbf{p}_1 - 1)$$

because we can move the constraint into the objective function with a Lagrange multiplier,  $\lambda$ . The maximum value must occur when the partial derivatives with respect to  $\mathbf{p}_1$ , our search variable, are zero:

$$\frac{\partial \phi}{\partial \mathbf{p}_1} = \frac{\partial(\mathbf{p}_1^T \mathbf{X}^T \mathbf{X} \mathbf{p}_1 - \lambda(\mathbf{p}_1^T \mathbf{p}_1 - 1))}{\partial \mathbf{p}_1} = 0$$

$$2\mathbf{X}^T \mathbf{X} \mathbf{p}_1 - 2\lambda_1 \mathbf{p}_1 = 0$$

$$(\mathbf{X}^T \mathbf{X} - \lambda_1 \mathbf{I}) \mathbf{p}_1 = 0$$

$$\mathbf{X}^T \mathbf{X} \mathbf{p}_1 = \lambda_1 \mathbf{p}_1$$

which is just the eigenvalue equation, indicating that  $\mathbf{p}_1$  is the eigenvector of  $\mathbf{X}^T \mathbf{X}$  and  $\lambda_1$  is the eigenvalue. One can show that  $\lambda_1 = \mathbf{t}_1^T \mathbf{t}_1$ , which is proportional to the variance of the first component. In a similar manner we can calculate the second eigenvalue, but this time we add the additional constraint that  $\mathbf{p}_1 \perp \mathbf{p}_2$ . Writing out this objective function and taking partial derivatives leads to showing that

$$\mathbf{X}^T \mathbf{X} \mathbf{p}_2 = \lambda_2 \mathbf{p}_2$$

From this we learn that: \* The loadings are the eigenvectors of  $\mathbf{X}^T \mathbf{X}$ . \* Sorting the eigenvalues in order from largest to smallest gives the order of the corresponding eigenvectors, the loadings. \* We know from the theory of eigenvalues that if there are distinct eigenvalues, then their eigenvectors are linearly independent (orthogonal). \* We also know the eigenvalues of  $\mathbf{X}^T \mathbf{X}$  must be real values and positive; this matches with the interpretation that the eigenvalues are proportional to the variance of each score vector. \* Also, the sum of the eigenvalues must add up to sum of the diagonal entries of  $\mathbf{X}^T \mathbf{X}$ , which represents of the total variance of the  $\mathbf{X}$  matrix, if all eigenvectors are extracted. So plotting the eigenvalues is equivalent to showing the proportion of variance explained in  $\mathbf{X}$  by each component. This is not necessarily a good way to judge the number of components to use, but it is a rough guide: use a Pareto plot of the eigenvalues (though in the context of eigenvalue problems, this plot is called a scree plot).

```
[8]: cov = np.cov(X, rowvar = False)
     evals , P = LA.eigh(cov)
     idx = np.argsort(evals)[::-1]
     P = P[:,idx]
     evals = evals[idx]
     T = np.dot(X, P)
```

```

Sigma=LA.norm(T,axis=0)
print("T (Scores)")
print(T)
print("P (Loadings)")
print(P)
print("Sigma (Variance)")
print(Sigma)

```

```

T (Scores)
[[ 4.25324997e+03  8.41288672e-01  8.37858943e-03]
 [-2.97275001e+03  1.25977271e-01 -1.82476780e-01]
 [-3.62475003e+03  1.56843494e-01  1.65224286e-01]
 [ 2.34425007e+03 -1.12410944e+00  8.87390454e-03]]
P (Loadings)
[[-1.21901390e-05 -5.66460727e-01  8.24088736e-01]
 [-9.99999997e-01 -5.32639789e-05 -5.14047691e-05]
 [-7.30130279e-05  8.24088734e-01  5.66460725e-01]]
Sigma (Variance)
[6.74994067e+03  1.41840009e+00  2.46466604e-01]

```

### 6.1 Task 1: Test if the loading vectors are orthogonal and orthonormal or not

[ ]:

### 6.2 Task 2: Test if the scores vectors are orthogonal and orthonormal or not

[ ]:

# PCA-Clustering-Excercise

February 17, 2021

## 1 Objective: Learn to do clustering and noise reduction in data using PCA

```
[1]: import matplotlib.pyplot as plt
import numpy as np
from numpy.linalg import svd
from sklearn.datasets import load_digits
digits = load_digits()
digits.data.shape
```

```
[1]: (1797, 64)
```

### 1.1 PCA using SVD

```
[2]: def pca(X):
    U, S, PTrans = svd(X, full_matrices=False)
    Sigma = np.diag(S)
    T=np.dot(U,Sigma)
    P=PTrans.T
    return T, Sigma, P #Score, Variace, Loadings
```

```
[3]: def plot_digits(data):
    fig, axes = plt.subplots(4, 10, figsize=(10, 4),
                             subplot_kw={'xticks':[], 'yticks':[]},
                             gridspec_kw=dict(hspace=0.1, wspace=0.1))
    for i, ax in enumerate(axes.flat):
        ax.imshow(data[i].reshape(8, 8),
                  cmap='binary', interpolation='nearest',
                  clim=(0, 16))
```

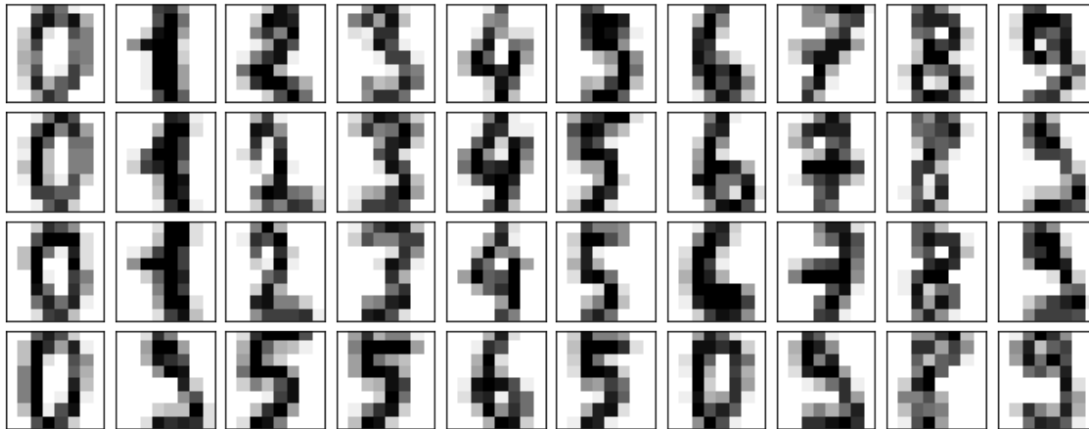
```
[4]: # Find out the original dimension of the data
X=digits.data
y=digits.target
print("Shape of X",X.shape)
print("Shape of y", y.shape)
```



Shape of X (1797, 64)

Shape of y (1797,)

```
[5]: #Visualize the original data
plot_digits(X)
```



**1.1.1 Task 1: Dimensionality reduction: Conduct PCA on the the matrix X to find out the dimension required to capture 80% of the variance**

```
[ ]: #SS?
explained_variance = (SS ** 2) / 4
explained_variance_ratio = (explained_variance / explained_variance.sum())
plt.plot(np.cumsum(explained_variance_ratio))
plt.grid()
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');
```

**1.1.2 Task 2: Clustering: Project the original data matrix X on the first two PCs and draw the scalar plot**

```
[ ]: #t1=?
#t2=?
plt.figure(figsize=(15,10))
plt.scatter(t1, t2,
            c=digits.target, edgecolor='none', alpha=0.5,
            cmap=plt.cm.get_cmap('Spectral', 10))
plt.xlabel('component 1')
plt.ylabel('component 2')
plt.colorbar();
```

### 1.1.3 Task 3: Denoising: Remove noise from the noisy data

```
[ ]: # Adding noise to the original data
X=digits.data
y=digits.target
np.random.seed(42)
noisy = np.random.normal(X, 4)
plot_digits(noisy)
```

Tips:

- Decompose the noisy data using PCA
- Reconstruct the data using just a few dominant components. For eg. check the variance plot

Since the nature of the noise is more or less similar across all the digits, they are not the features with enough variance to discriminate between the digits.

```
[ ]: #X_denoised = ?
plot_digits(X_denoised)
```

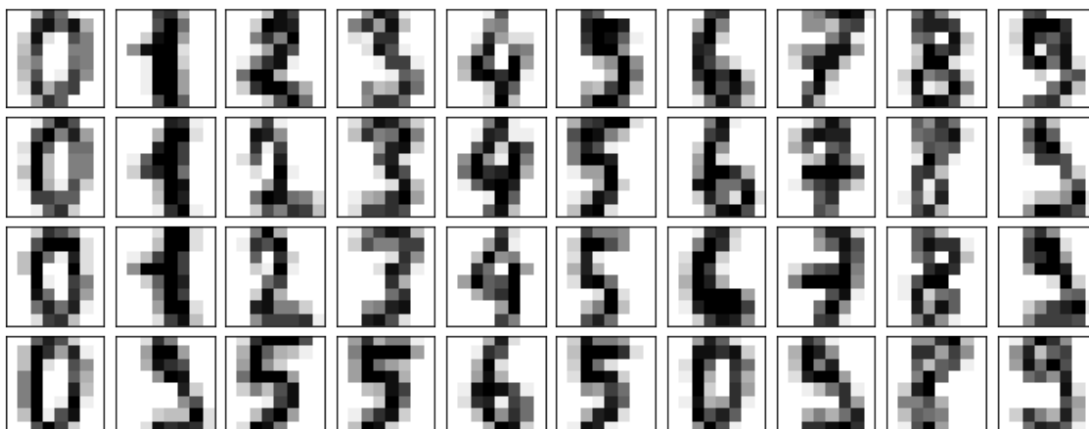
## 2 All the above exercise can be done using the SKLEARN library as follows

```
[6]: from sklearn.decomposition import PCA
X=digits.data
y=digits.target

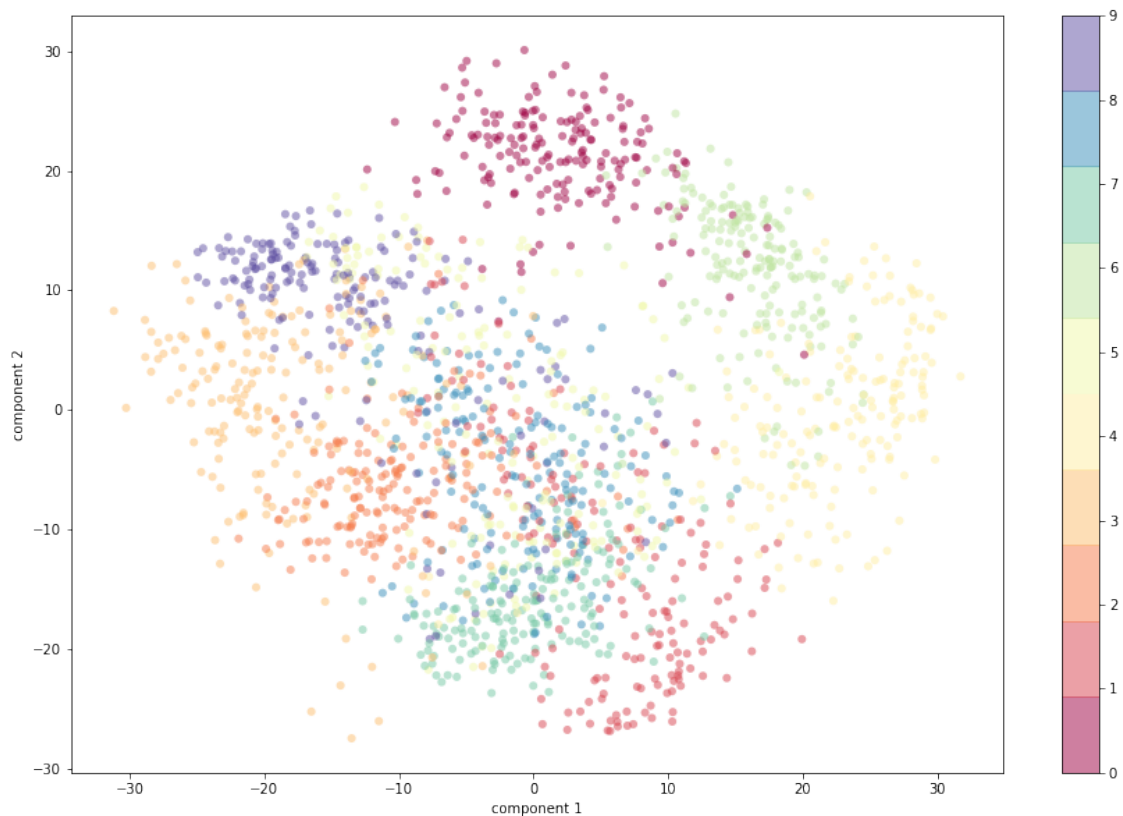
[7]: pca = PCA(2) # project from 64 to 2 dimensions
projected = pca.fit_transform(digits.data)
print(digits.data.shape)
print(projected.shape)
plot_digits(digits.data)
```

(1797, 64)

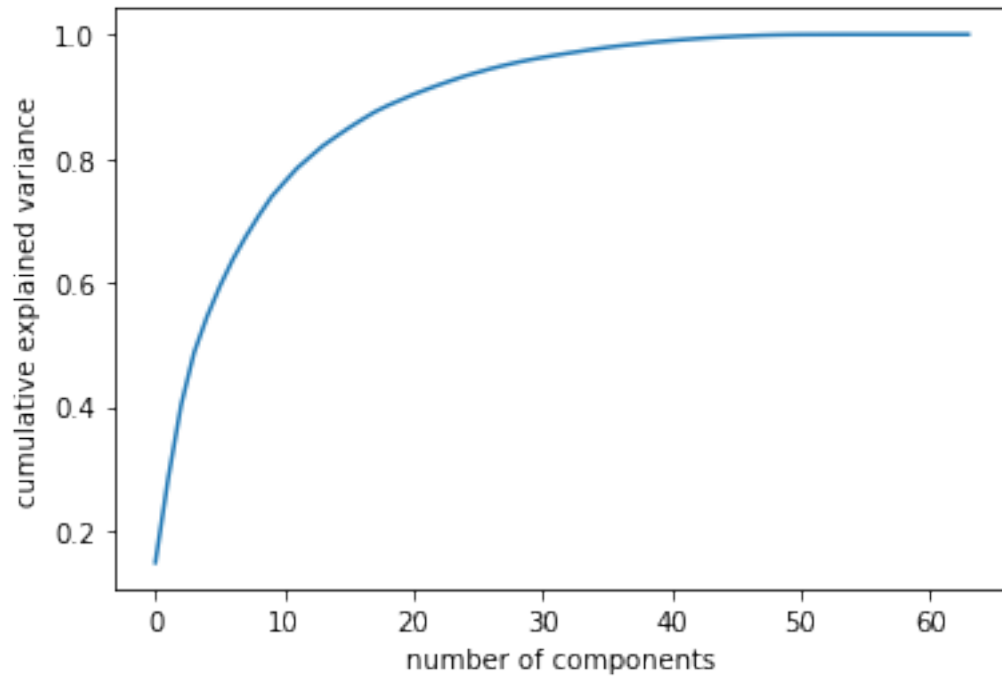
(1797, 2)



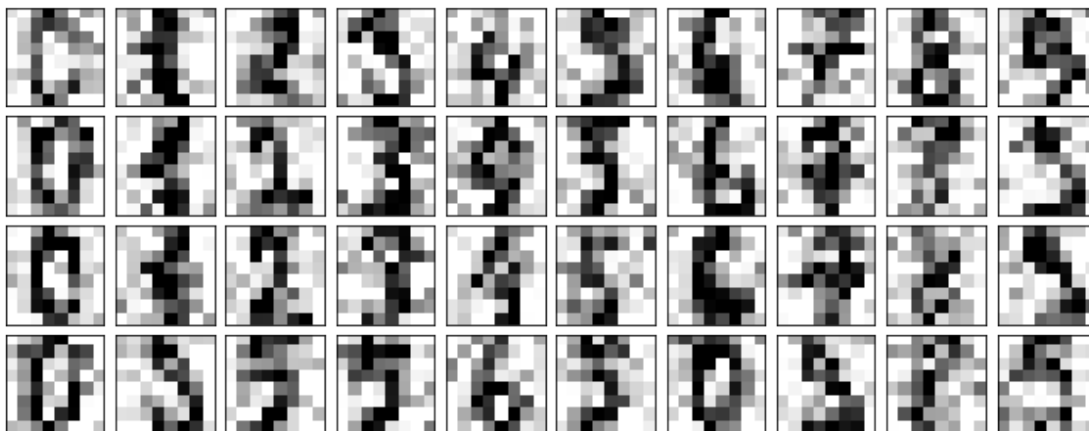
```
[8]: plt.figure(figsize=(15,10))
plt.scatter(projected[:, 0], projected[:, 1],
            c=digits.target, edgecolor='none', alpha=0.5,
            cmap=plt.cm.get_cmap('Spectral', 10))
plt.xlabel('component 1')
plt.ylabel('component 2')
plt.colorbar();
```



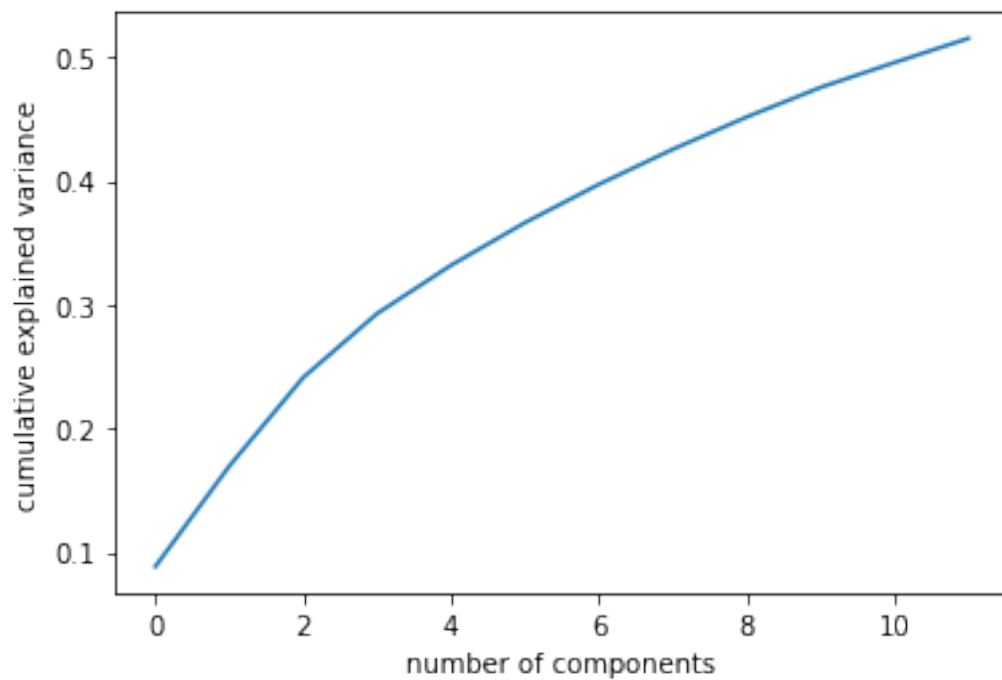
```
[9]: pca = PCA().fit(digits.data)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');
```



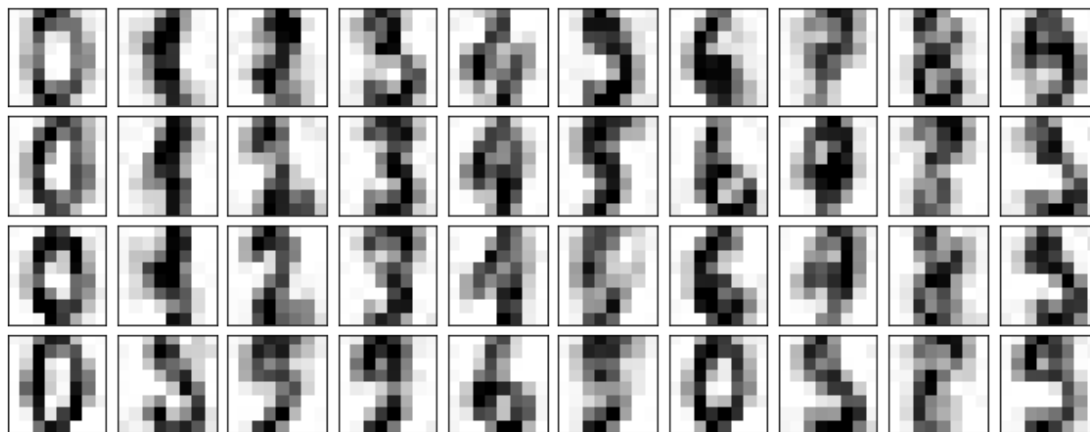
```
[10]: np.random.seed(42)
      noisy = np.random.normal(digits.data, 4)
      plot_digits(noisy)
```



```
[11]: pca = PCA(0.50).fit(noisy) # 50% of the variance amounts to 12 principal
      → components.
      pca.n_components_
      plt.plot(np.cumsum(pca.explained_variance_ratio_))
      plt.xlabel('number of components')
      plt.ylabel('cumulative explained variance');
```



```
[12]: components = pca.transform(noisy)
      filtered = pca.inverse_transform(components)
      plot_digits(filtered)
```



# Scaling-on-PCA-Excercise

February 17, 2021

## 1 Objective: To understand the importance of scaling on PCA

```
[60]: from sklearn.decomposition import PCA
      from sklearn import preprocessing
      from sklearn import metrics
      import numpy as np
      import matplotlib.pyplot as plt
      from sklearn.datasets import load_wine
```

## 2 Task 0: Write the function to compute the pca using Eigenvector approach

```
[61]: from numpy.linalg import svd
      def pca(X):
          #Populate the code
          return T, Sigma, P #Score, Variance, Loadings
```

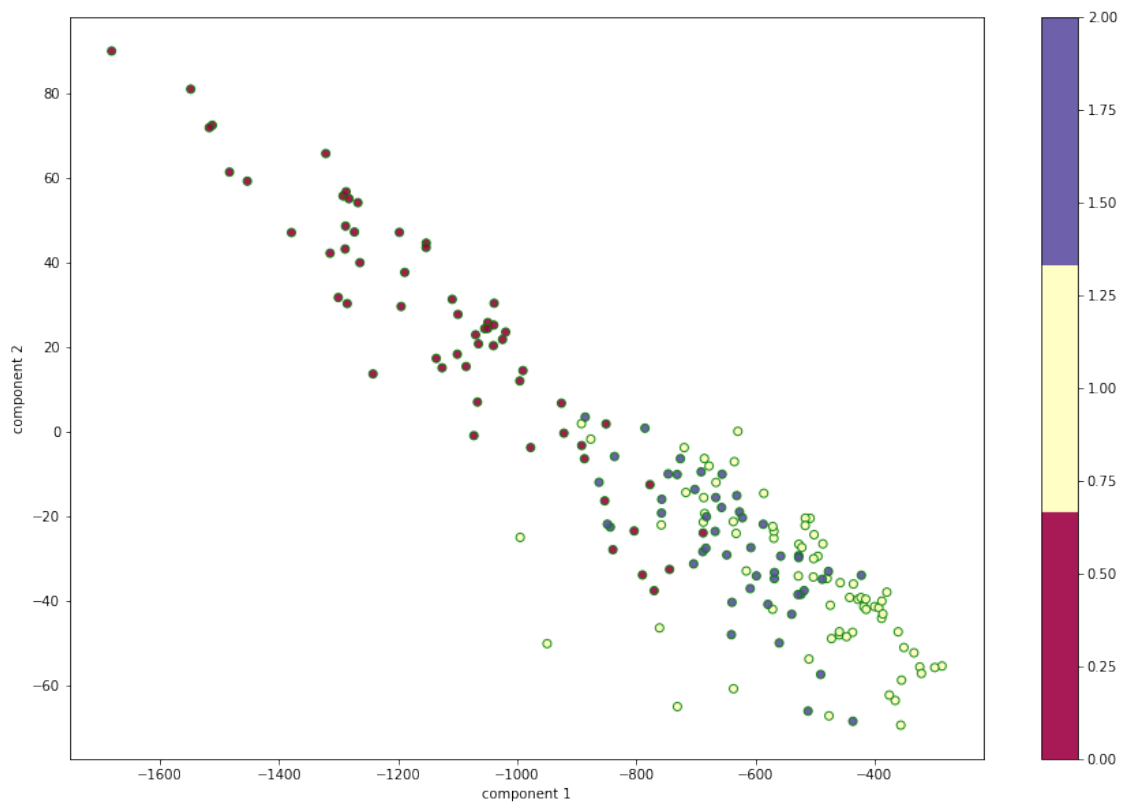
```
[62]: features, target = load_wine(return_X_y=True)
      X=features
      y=target
```

## 3 Three different ways of scaling

- Scaling by removing the mean and dividing by the standard deviation `#standard_scaling=preprocessing.StandardScaler()`  
`#X_standard=standard_scaling.fit_transform(X)`
- Scaling to min and maximum values of each feature `#min-max_scaling=preprocessing.MinMaxScaler()`  
`#X_minmax=minmax_scaling.fit_transform(X)`
- Scaling by dividing by the maximum absolute values of each features `#max_abs_scaler=preprocessing.MaxAbsScaler()`  
`#X_maxabs=max_abs_scaler.fit_transform(X)`

### 3.1 Task 1: Create the scores plot without any scaling

```
[63]: T,S,P=pca(X)
plt.figure(figsize=(15,10))
plt.scatter(T[:, 0], T[:, 1],
            c=y, edgecolor='green', alpha=0.9,
            cmap=plt.cm.get_cmap('Spectral', 3))
plt.xlabel('component 1')
plt.ylabel('component 2')
plt.colorbar();
```



### 3.2 Task 2: Create the scores plot without any standard scaling

```
[ ]:
```

### 3.3 Task 3: Create the scores plot without any min max scaling

```
[ ]:
```

### 3.4 Task 4: Create the scores plot without any max abs scaling

[ ]:



# WindSolarPower

February 17, 2021

```
[15]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[16]: x_path = '../data/ElectricPowerProduction/X.txt'
y_path = '../data/ElectricPowerProduction/Y.txt'
```

```
[17]: x=pd.read_csv(x_path,sep=';')
y=pd.read_csv(y_path,sep=';')
```

```
[18]: x.head()
```

```
[18]:
```

	Year	IsDayBin	Month	WindSpeed	IrrDirect	IrrDiffuse	Temperature	\
R1	2016	Night	1	5.777	0.0	0.0	-1.210	
R2	2016	Night	1	5.887	0.0	0.0	-0.862	
R3	2016	Night	1	5.957	0.0	0.0	-0.552	
R4	2016	Night	1	6.025	0.0	0.0	-0.297	
R5	2016	Night	1	6.101	0.0	0.0	-0.149	

	Percipitation	SnowFlow	SnowMass	AirDensity	RadSurface	RadTop	\
R1	0.114	0.099	10.592	1.268	0.0	0.0	
R2	0.114	0.103	10.671	1.266	0.0	0.0	
R3	0.122	0.111	10.756	1.265	0.0	0.0	
R4	0.158	0.145	10.855	1.264	0.0	0.0	
R5	0.164	0.081	10.884	1.263	0.0	0.0	

	CloudCover
R1	1.0
R2	1.0
R3	1.0
R4	1.0
R5	1.0

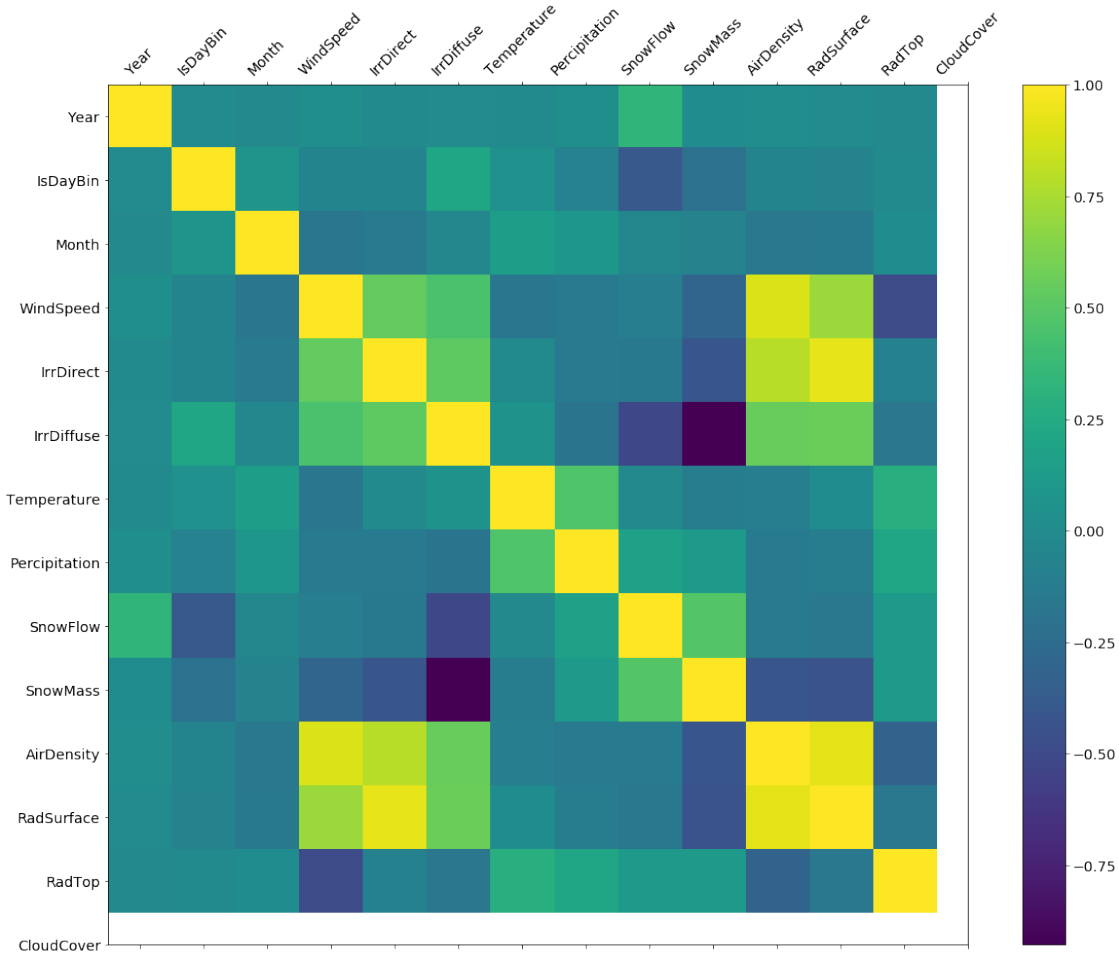
```
[19]: y.head()
```

```
[19]:
```

	WindPower	PVPower
R1	396.352	0.0
R2	417.778	0.0
R3	434.581	0.0
R4	448.482	0.0

R5      465.790      0.0

```
[20]: f = plt.figure(figsize=(19, 15))
plt.matshow(x.corr(), fignum=f.number)
plt.xticks(range(x.shape[1]), x.columns, fontsize=14, rotation=45)
plt.yticks(range(x.shape[1]), x.columns, fontsize=14)
cb = plt.colorbar()
cb.ax.tick_params(labelsize=14)
```



```
[21]: x['IsDayBin'].unique()
```

```
[21]: array(['Night', 'Day'], dtype=object)
```

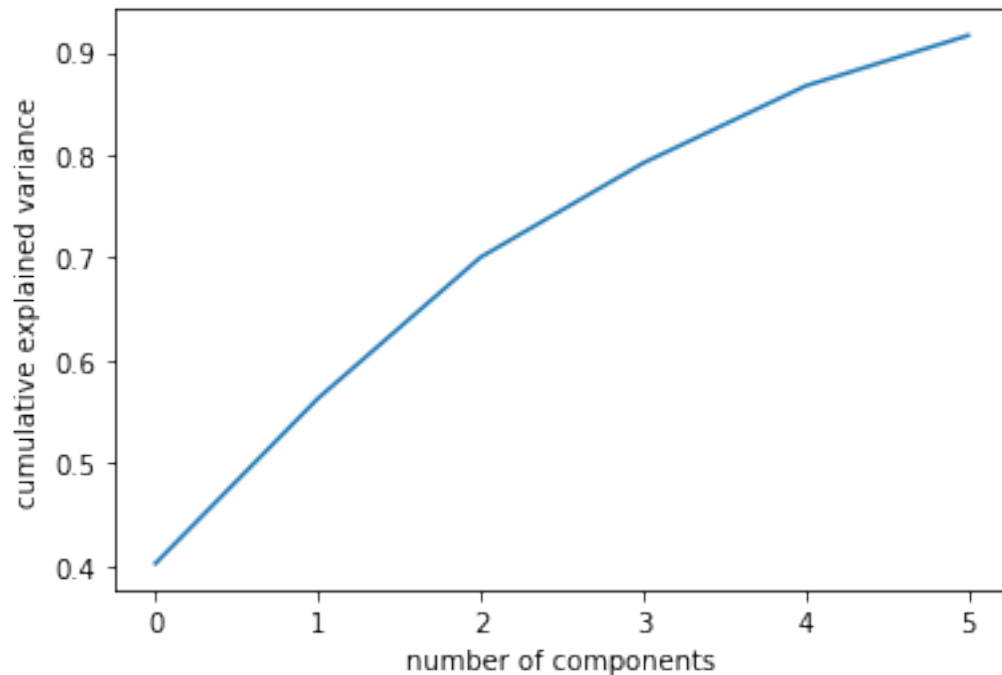
```
[22]: X=x.drop(['Year', 'IsDayBin', 'Month'],axis=1)
```

```
[23]: X=X.values
Y=y.values
```

```
[24]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
```

```
scaler.fit(X)
X_scaled=scaler.transform(X)
```

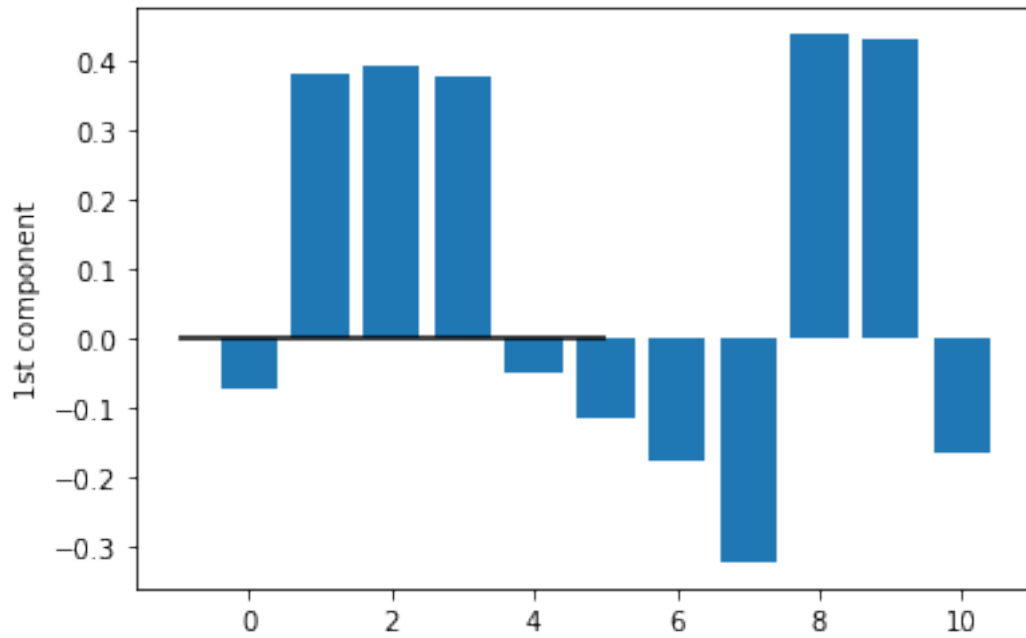
```
[25]: from sklearn.decomposition import PCA
pca = PCA(6).fit(X_scaled)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');
```



```
[27]: P=pca.components_
T = X_scaled.dot(P.T)
```

```
[28]: #WindSpeed      IrrDirect      IrrDiffuse      Temperature      Percipitation
print("First component ", P.T[:,0])
x = np.arange(X_scaled.shape[1])
plt.bar(x, height= P.T[:,0])
plt.hlines(0,-1,5)
plt.ylabel("1st component")
plt.show()
```

```
First component  [-0.07338616  0.38030048  0.39169996  0.37810002 -0.04815644
-0.11656518
-0.17705383 -0.32159164  0.43704876  0.42920331 -0.16641727]
```



## 1 Task 0: Explore the dataset using what you have learned so far

[ ]: