Principal Component Analysis

Dr. Víctor Uc Cetina

Facultad de Matemáticas
Universidad Autónoma de Yucatán

cetina@informatik.uni-hamburg.de
https://sites.google.com/view/victoruccetina

Content

Principal Component Analysis

- We will develop a method that tries to identify the subspace in which the data approximately lies: Principal Component Analysis (PCA).

Principal Component Analysis

- We will develop a method that tries to identify the subspace in which the data approximately lies: Principal Component Analysis (PCA).
- PCA requires only an eigenvector calculation and does not need to resort to EM.

Principal Component Analysis

- PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences.

Principal Component Analysis

- PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences.
- Since patterns in data can be hard to find in data of high dimension, where a graphical representation is not available, PCA is a powerful tool for analysing data.

Principal Component Analysis

- The other main advantage of PCA is that once you have found these patterns in the data, you can compress the data by reducing the number of dimensions, without much loss of information.

Principal Component Analysis

- The other main advantage of PCA is that once you have found these patterns in the data, you can compress the data by reducing the number of dimensions, without much loss of information.
- This technique is used in image compression, for example.

Principal Component Analysis

The steps for applying PCA are:

1. Get some data.

Principal Component Analysis

The steps for applying PCA are:

1. Get some data.

2. Substract the mean.

Principal Component Analysis

The steps for applying PCA are:

1. Get some data.
2. Substract the mean.
3. Calculate the covariance matrix.

Principal Component Analysis

The steps for applying PCA are:

1. Get some data.
2. Substract the mean.
3. Calculate the covariance matrix.
4. Calculate the eigenvectors and eigenvalues of the covariance matrix.

Principal Component Analysis

The steps for applying PCA are:

1. Get some data.
2. Substract the mean.
3. Calculate the covariance matrix.
4. Calculate the eigenvectors and eigenvalues of the covariance matrix.
5. Choose components and form a feature vector.

Principal Component Analysis

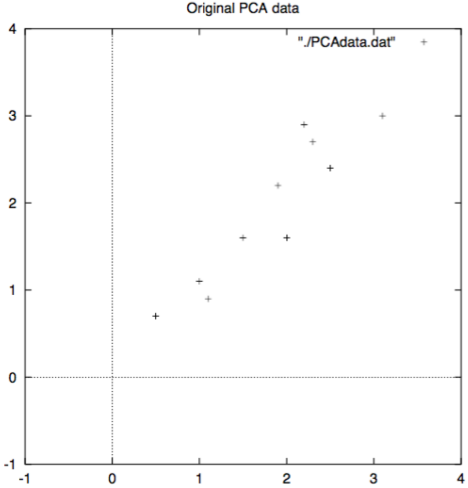
The steps for applying PCA are:

1. Get some data.
2. Substract the mean.
3. Calculate the covariance matrix.
4. Calculate the eigenvectors and eigenvalues of the covariance matrix.
5. Choose components and form a feature vector.
6. Derive the new data set.

1. Get Some Data.

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9



2. Substract the Mean

- For PCA to work properly, you have to subtract the mean from each of the data dimensions.

2. Substract the Mean

- For PCA to work properly, you have to subtract the mean from each of the data dimensions.
- The mean subtracted is the average across each dimension.

2. Substract the Mean

- For PCA to work properly, you have to subtract the mean from each of the data dimensions.
- The mean subtracted is the average across each dimension.
- So, all the x_1 values have μ_1 (the mean of the x_1 values of all the data points) subtracted, and all the values x_2 have μ_2 subtracted from them.

2. Substract the Mean

- This produces a data set whose mean is zero.

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01

3. Calculate the Covariance Matrix

- The covariance matrix for this example is

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

3. Calculate the Covariance Matrix

- The covariance matrix for this example is

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

- So, since the non-diagonal elements in this covariance matrix are positive, we should expect that both the x_1 and x_2 variable increase together.

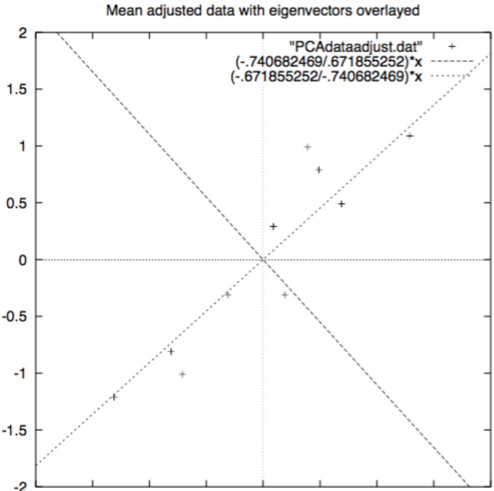
4. Calculate the Eigenvectors and Eigenvalues of the Covariance Matrix

- Since the covariance matrix is square, we can calculate the eigenvectors and eigenvalues for this matrix. These are rather important, as they tell us useful information about our data.

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

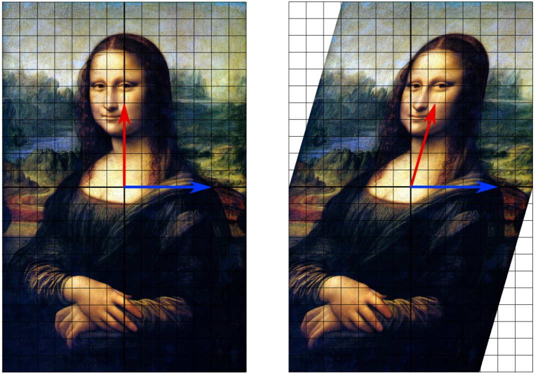
$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

4. Calculate the Eigenvectors and Eigenvalues of the Covariance Matrix



4. Calculate the Eigenvectors and Eigenvalues of the Covariance Matrix

A (non-zero) vector v of dimension N is an eigenvector of a square $N \times N$ matrix A if it satisfies the linear equation $A\mathrm{v} = \lambda\mathrm{v}$, where λ is a scalar and the eigenvalue corresponding to v.



5. Choosing Components and Forming a Feature Vector

- It turns out that the eigenvector with the highest eigenvalue is the principle component of the data set.

5. Choosing Components and Forming a Feature Vector

- It turns out that the eigenvector with the highest eigenvalue is the principle component of the data set.
- In our example, the eigenvector with the largest eigenvalue was the one that pointed down the middle of the data.

5. Choosing Components and Forming a Feature Vector

- It turns out that the eigenvector with the highest eigenvalue is the principle component of the data set.
- In our example, the eigenvector with the largest eigenvalue was the one that pointed down the middle of the data.
- It is the most significant relationship between the data dimensions.

5. Choosing Components and Forming a Feature Vector

- In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest.

5. Choosing Components and Forming a Feature Vector

- In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest.
- This gives you the components in order of significance.

5. Choosing Components and Forming a Feature Vector

- In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalue, highest to lowest.
- This gives you the components in order of significance.
- If you like, you can decide to ignore the components of lesser significance. You do lose some information, but if the eigenvalues are small, you don't lose much.

6. Deriving the New Data Set

- This the final step in PCA, and is also the easiest.

6. Deriving the New Data Set

- This the final step in PCA, and is also the easiest.
- Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

6. Deriving the New Data Set

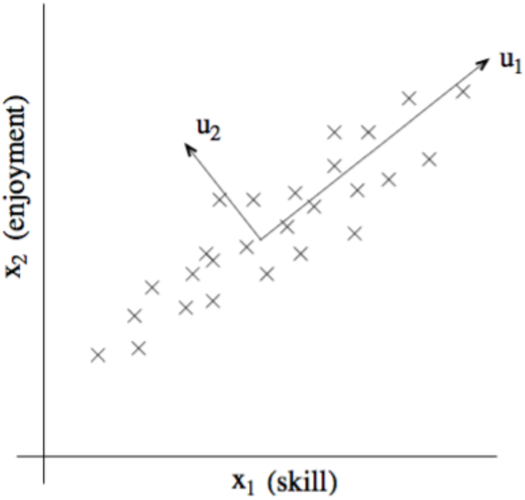
- This the final step in PCA, and is also the easiest.
- Once we have chosen the components (eigenvectors) that we wish to keep in our data and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.
- It will give us the original data solely in terms of the vectors we chose.

Principal Component Analysis

Prior to running PCA, we first pre-process the data to normalize its mean and variance as follows:

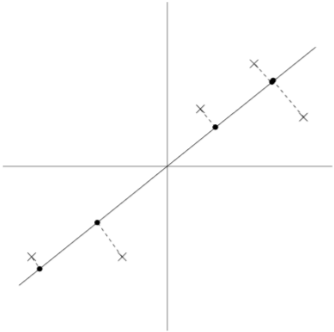
1. Let $\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)}$.
2. Replace each $x^{(i)}$ with $x^{(i)} - \mu$.
3. Let $\sigma_j^2 = \frac{1}{m}\sum_i (x_j^{(i)})^2$
4. Replace each $x_j^{(i)}$ with $x_j^{(i)}/\sigma_j$.

Principal Component Analysis



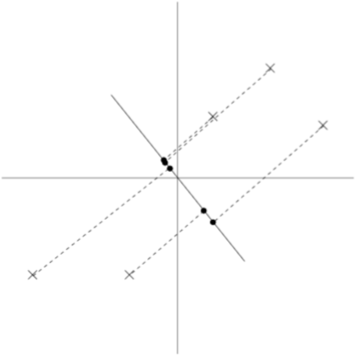
Principal Component Analysis

First principal component



Principal Component Analysis

Second principal component



Principal Component Analysis

	England	Wales	Scotland	N Ireland
Cheese	105	103	103	66
Carcass meat	245	227	242	267
Other meat	685	803	750	586
Fish	147	160	122	93
Fats and oils	193	235	184	209
Sugars	156	175	147	139
Fresh potatoes	720	874	566	1033
Fresh Veg	253	265	171	143
Other Veg	488	570	418	355
Processed potatoes	198	203	220	187
Processed Veg	360	365	337	334
Fresh fruit	1102	1137	957	674
Cereals	1472	1582	1462	1494
Beverages	57	73	53	47
Soft drinks	1374	1256	1572	1506
Alcoholic drinks	375	475	458	135
Confectionery	54	64	62	41

Table 1: UK food consumption in 1997 (g/person/week). Source: DEFRA website

Principal Component Analysis

Figure 1: Projections onto first principal component (1-D space)

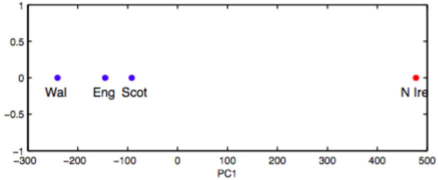
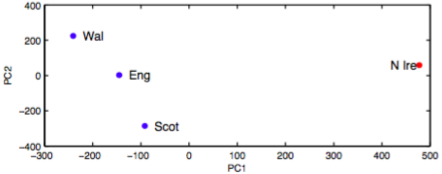


Figure 2: Projections onto first 2 principal components (2-D space)

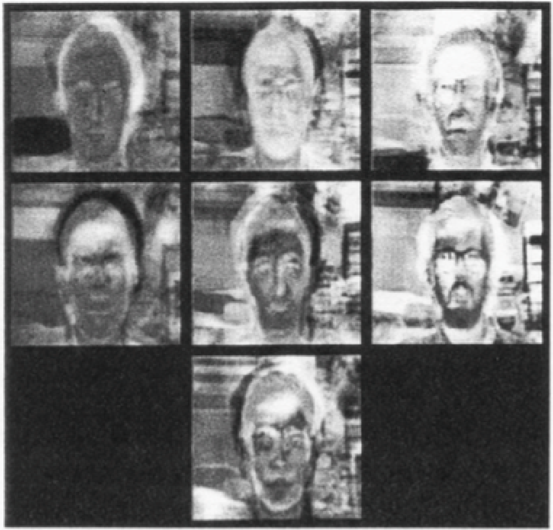


Face Recognition

Matthew Turk and Alex Pentland

Face Recognition

Face Recognition



Face Recognition

Eigenfaces Method(1/2)

1. Acquire an initial set of face images (the training set).

Eigenfaces Method(1/2)

1. Acquire an initial set of face images (the training set).

2. Calculate the eigenfaces from the training set, keeping only the M images that correspond to the highest eigenvalues. These M images define the *face space*. As new faces are experienced, the eigenfaces can be updated or recalculated.

Eigenfaces Method(1/2)

1. Acquire an initial set of face images (the training set).
2. Calculate the eigenfaces from the training set, keeping only the M images that correspond to the highest eigenvalues. These M images define the *face space*. As new faces are experienced, the eigenfaces can be updated or recalculated.
3. Calculate the corresponding distribution in M-dimensional weight space for each known individual, by projecting their face images onto the *face space*.

Eigenfaces Method (2/2)

1. Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.

Eigenfaces Method (2/2)

1. Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.

2. Determine if the image is a face at all by checking to see if the image is sufficiently close to *face space*.

Eigenfaces Method (2/2)

1. Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.

2. Determine if the image is a face at all by checking to see if the image is sufficiently close to *face space*.

3. If it is a face, classify the weight pattern as either, a known person or as unknown.

Eigenfaces Method (2/2)

1. Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.
2. Determine if the image is a face at all by checking to see if the image is sufficiently close to *face space*.
3. If it is a face, classify the weight pattern as either, a known person or as unknown.
4. (Optional) Update the eigenfaces and/or weight patterns.

Eigenfaces Method (2/2)

1. Calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.

2. Determine if the image is a face at all by checking to see if the image is sufficiently close to *face space*.

3. If it is a face, classify the weight pattern as either, a known person or as unknown.

4. (Optional) Update the eigenfaces and/or weight patterns.

5. (Optional) If the same unknown face is seen several times, calculate its characteristic weight pattern and incorporate it into the known faces.

Reference

- Matthew Turk, Alex Pentland. **Eigenfaces for recognition**. Journal of Cognitive Neuroscience. 3 (1): 71-86. 1991.
- Andrew Ng. **Machine Learning Course Notes**. 2003.

Thank you!

Dr. Victor Uc Cetina
cetina@informatik.uni-hamburg.de