Quality Data Analysis

Principal Component Analysis – Multivariate data

Bianca Maria Colosimo – biancamaria.colosimo@polimi.it

Principal Component Analysis – I.T. Joliffe- 2nd edition - Springer

http://web.stanford.edu/class/stats202/ - Data Mining and Analysis

Principal Component Analysis

This is the most popular unsupervised procedure ever.

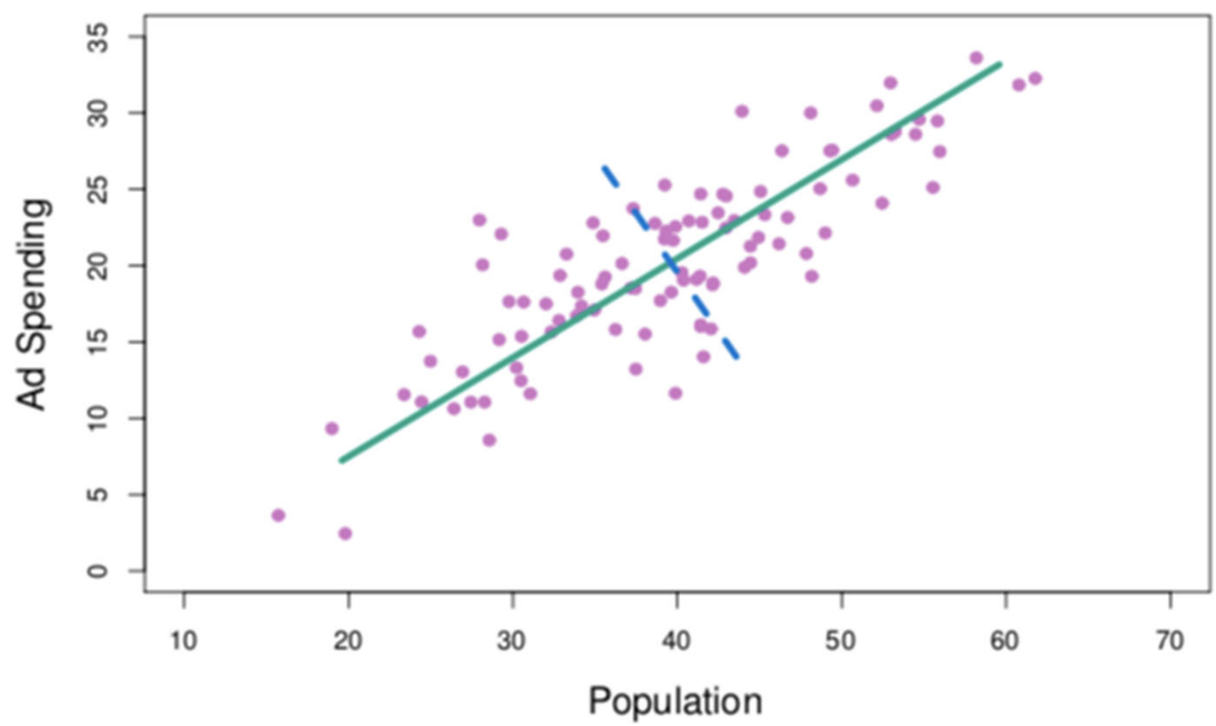
- Invented by Karl Pearson (1901).
- Developed by Harold Hotelling (1933).
- Also known as the Karhunen-Loeve Transform (KLT) (or Hotelling Transform and Eigenvector Transform),

It provides a way to visualize high dimensional data, summarizing the most important information.

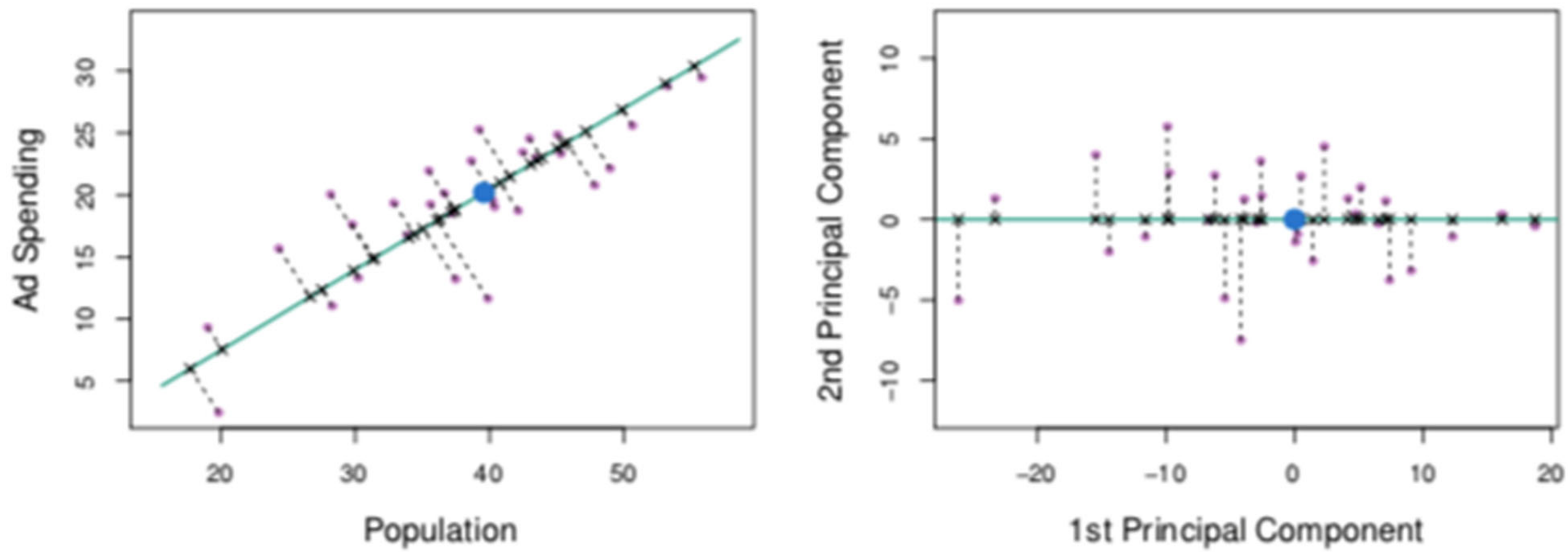
PCA

Principal components are a sequence of projections of the data, mutually uncorrelated and ordered in variance.

The first principal component is the direction of the line that minimizes the total squared distance from each point to its orthogonal projection onto the line.



1st and 2nd PCs



What does it happen with three variables?

The first two principal components span a plane which is closest to the data.

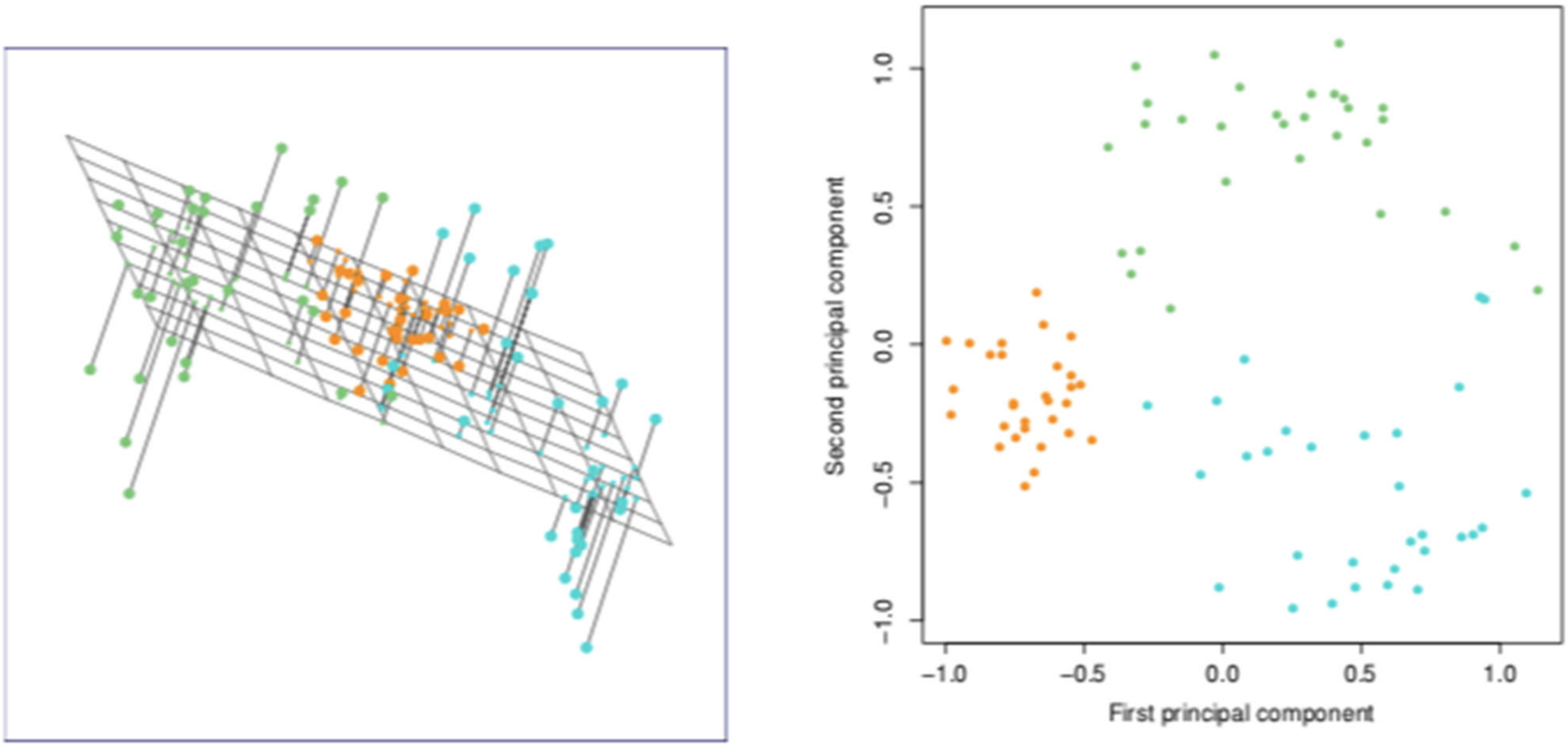
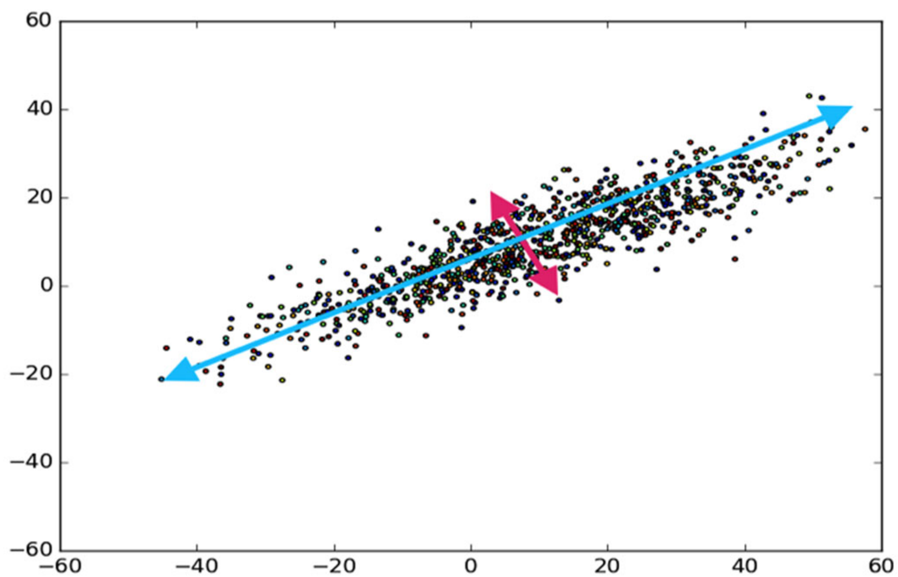


Figure 10.2

Assume we have two random variables

Can we find a new reference system that allows one to summarize all the variability observed in the two random variables using just one variable?



Multivariate random variables

Consider a p-component vector, i.e. a vector of p random variables

$$\boldsymbol{x}' = [x_1, x_2, \ldots, x_p]$$

Expected value $\boldsymbol{\mu}' = E(\boldsymbol{x}) = [E(x_1), E(x_2), \ldots, E(x_p)] = [\mu_1, \mu_2, \ldots, \mu_p]$

Variance-Covariance Matrix

$$\underset{pxp}{V(\boldsymbol{x})} = E[\underset{px1}{(\boldsymbol{x} - E(\boldsymbol{x}))}\underset{1xp}{(\boldsymbol{x} - E(\boldsymbol{x}))'}] = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{1p} & \sigma_{2p} & \ldots & \sigma_p^2 \end{bmatrix}$$

$$\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = \mathrm{cov}(x_i, x_j)$$

Remember

Product between a constant vector and a random vector:

Scalar (1xp)(px1)

$$V(\mathbf{a}'\mathbf{x}) = E[(\mathbf{a}'\mathbf{x} - E(\mathbf{a}'\mathbf{x}))(\mathbf{a}'\mathbf{x} - E(\mathbf{a}'\mathbf{x}))'] =$$
$$= E[(\mathbf{a}'\mathbf{x} - \mathbf{a}'E(\mathbf{x}))(\mathbf{a}'\mathbf{x} - \mathbf{a}'E(\mathbf{x}))'] = \mathbf{a}'E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{a}'(\mathbf{x} - E(\mathbf{x})))'] =$$
$$= \mathbf{a}'E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))']\mathbf{a} = \mathbf{a}'\mathbf{\Sigma}\mathbf{a}$$

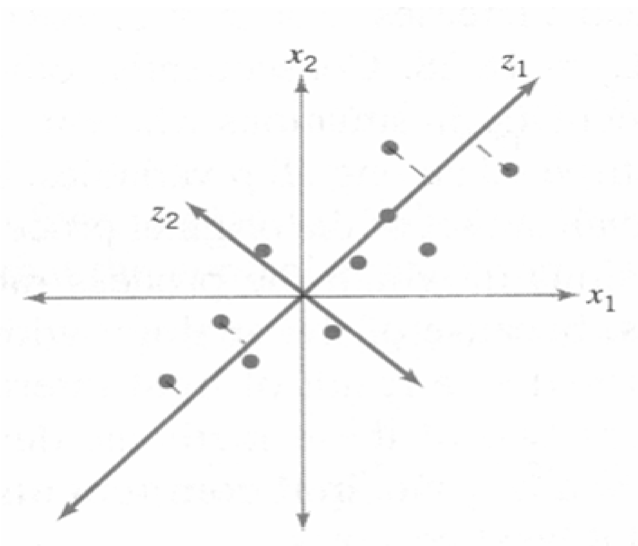
correlation

$$\rho_{ij} = \frac{\operatorname{cov}(x_i x_j)}{\sqrt{V(x_i)V(x_j)}} \qquad \boldsymbol{P} = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & 1 & \dots & \rho_{2p} \\ \dots & \dots & \dots & \dots \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{bmatrix}$$

Correlation matrix

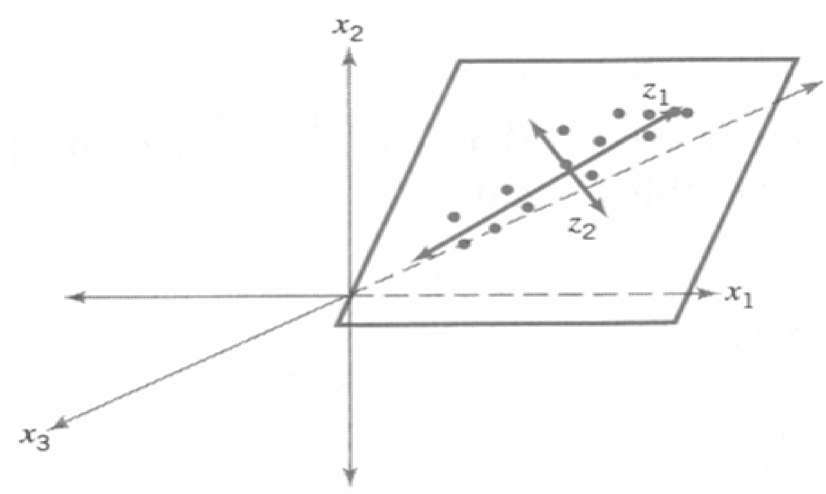
Given a set of *p* process variables, the goal is to perform a reference system transformation:

$$x_1, x_2, \ldots, x_p \longrightarrow z_1, z_2, \ldots, z_p$$



(a) $p = 2$

z_1: largest percentage of data variability



(b) $p = 3$

PCA

The key idea consists of finding a new orthonormal reference system that allows maximizing the variability of original data and, simultaneously, reducing the number of variables that are necessary to describe the process

$$
\begin{cases}
z_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \boldsymbol{\alpha'}_1\boldsymbol{x} \\
z_2 = \alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2p}x_p = \boldsymbol{\alpha'}_2\boldsymbol{x} \\
\\
z_p = \alpha_{p1}x_1 + \alpha_{p2}x_2 + \cdots + \alpha_{pp}x_p = \boldsymbol{\alpha'}_P\boldsymbol{x}
\end{cases}
$$

PCA

The first step is to look for a linear function of the elements of **x** having maximum variance

$$\boldsymbol{\alpha}_1'\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j.$$

Next, look for a linear function $\alpha_2{}'\mathbf{x}$
uncorrelated with $\alpha_1{}'\mathbf{x}$ having maximum variance, and so on…

At the k-th stage, look for a linear function that has maximum variance being uncorrelated with all the previous linear combinations

Joliffe – Principal Component Analysis - Introduction

PCA

Let Σ be the covariance matrix of **X**

Let **X** be a data matrix with n samples, and *p* variables. From each variable, we subtract the mean of the column; i.e. we center the variables.

We can found up to *p* PCs, but in general, most of the variability in **X** can be accounted by *m* PCs, where m<<p

Usually Σ is unknown and can be replaced by the sample covariance matrix **S**

Consider $\boldsymbol{\alpha}'_1\boldsymbol{x}$. The vector $\boldsymbol{\alpha}'_1$ maximizes var[$\boldsymbol{\alpha}'_1\boldsymbol{x}$]= $\boldsymbol{\alpha}'_1\Sigma\,\boldsymbol{\alpha}_1$ normalization constraint $\boldsymbol{\alpha}'_1\boldsymbol{\alpha}_1$=1 sum of squares of elements of $\boldsymbol{\alpha}_1$ equals 1.

PCA

Constrained optimization: Lagrange multipliers

Maximize $\quad \boldsymbol{\alpha}'_1 \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1 - \lambda(\boldsymbol{\alpha}'_1 \boldsymbol{\alpha}_1 - 1)$

Differentiate with respect to $\boldsymbol{\alpha}_1$ gives $\quad \boldsymbol{\Sigma}\boldsymbol{\alpha}_1 - \lambda\boldsymbol{\alpha}_1 = \mathbf{0}$

$$(\Sigma - \lambda I_p)\boldsymbol{\alpha}_1 = \mathbf{0}$$

Thus λ is an eigenvalue of Σ and $\boldsymbol{\alpha}_1$ is the corresponding eigenvector.
Note that the quantity to be maximized is the variance, given by

$$\boldsymbol{\alpha}'_1 \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}'_1 \lambda\, \boldsymbol{\alpha}_1 = \lambda \boldsymbol{\alpha}'_1\, \boldsymbol{\alpha}_1 = \lambda$$

Thus $\boldsymbol{\alpha}_1$ is the eigenvector corresponding to the largest eigenvalue of Σ and $\mathrm{var}(\boldsymbol{\alpha}'_1 \boldsymbol{x}) = \boldsymbol{\alpha}'_1 \boldsymbol{\Sigma}\, \boldsymbol{\alpha}_1 = \lambda_1$ the largest eigenvalue.

The second PC, $\boldsymbol{\alpha}'_2\boldsymbol{x}$ maximizes $\boldsymbol{\alpha}'_2\boldsymbol{\Sigma}\boldsymbol{\alpha}_2$ s.t. being uncorrelated with $\boldsymbol{\alpha}'_2\boldsymbol{x}$ i.e., cov[$\boldsymbol{\alpha}'_1\boldsymbol{x}$, $\boldsymbol{\alpha}'_2\boldsymbol{x}$]=0

cov[$\boldsymbol{\alpha}'_1\boldsymbol{x}$, $\boldsymbol{\alpha}'_2\boldsymbol{x}$]= $\boldsymbol{\alpha}'_1\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}'_2\boldsymbol{\Sigma}\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}'_2\lambda_1\boldsymbol{\alpha}_1 = \lambda_1\boldsymbol{\alpha}'_2\boldsymbol{\alpha}_1 = \lambda_1\boldsymbol{\alpha}'_1\boldsymbol{\alpha}_2 = 0$

$$\boxed{\boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda(\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_2 - 1) - \phi\boldsymbol{\alpha}_2'\boldsymbol{\alpha}_1}$$

where λ, ϕ are Lagrange multipliers. Differentiation with respect to $\boldsymbol{\alpha}_2$ gives

$$\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_2 - \phi\boldsymbol{\alpha}_1 = \mathbf{0}$$

$$\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2 - \phi\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 0,$$

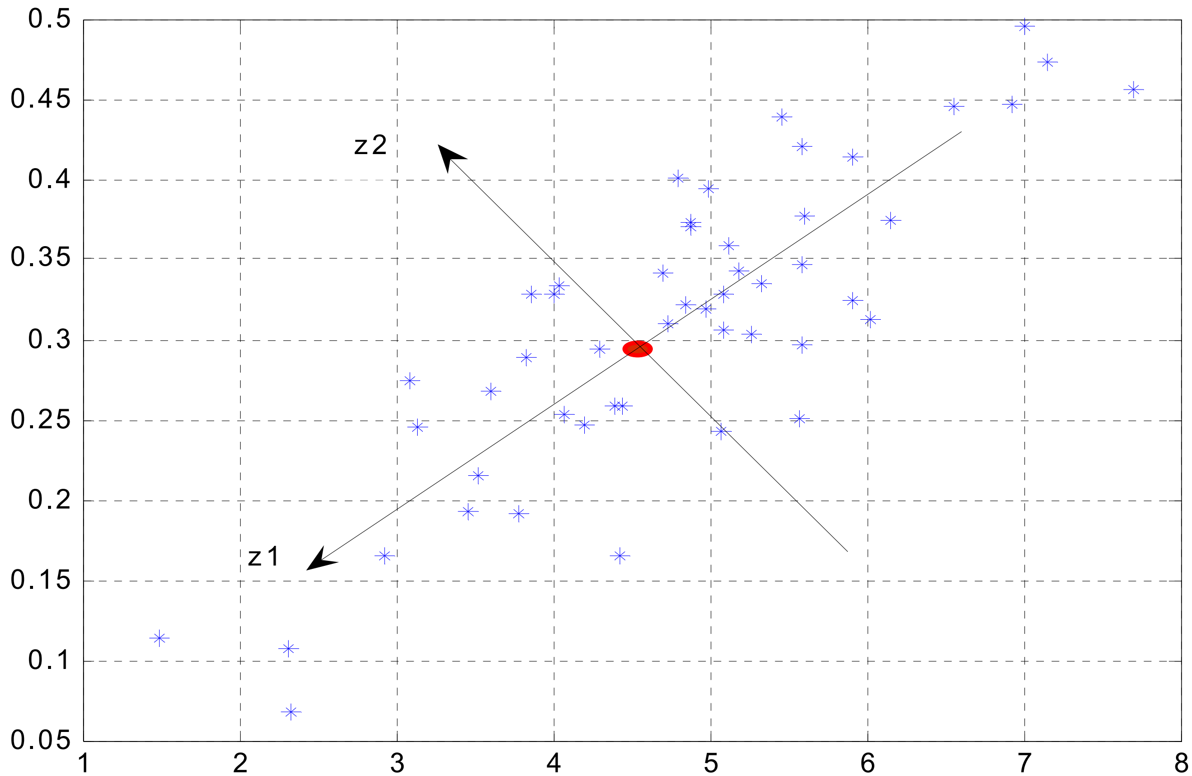
which, since the first two terms are zero and $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_1 = 1$, reduces to $\phi = 0$. Therefore, $\boldsymbol{\Sigma}\boldsymbol{\alpha}_2 - \lambda\boldsymbol{\alpha}_2 = \mathbf{0}$, or equivalently $(\boldsymbol{\Sigma} - \lambda\mathbf{I}_p)\boldsymbol{\alpha}_2 = \mathbf{0}$, so λ is once more an eigenvalue of $\boldsymbol{\Sigma}$, and $\boldsymbol{\alpha}_2$ the corresponding eigenvector.

Again, $\lambda = \boldsymbol{\alpha}_2'\boldsymbol{\Sigma}\boldsymbol{\alpha}_2$, so λ is to be as large as possible. Assuming that $\boldsymbol{\Sigma}$ does not have repeated eigenvalues, a complication that is discussed in Section 2.4, λ cannot equal λ_1. If it did, it follows that $\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1$, violating the constraint $\boldsymbol{\alpha}_1'\boldsymbol{\alpha}_2 = 0$. Hence λ is the second largest eigenvalue of $\boldsymbol{\Sigma}$, and $\boldsymbol{\alpha}_2$ is the corresponding eigenvector.

As stated above, it can be shown that for the third, fourth, $\dots$, pth PCs, the vectors of coefficients $\boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4, \dots, \boldsymbol{\alpha}_p$ are the eigenvectors of $\boldsymbol{\Sigma}$ corresponding to $\lambda_3, \lambda_4, \dots, \lambda_p$, the third and fourth largest, $\dots$, and the smallest eigenvalue, respectively. Furthermore,

$$\mathrm{var}[\boldsymbol{\alpha}_k'\mathbf{x}] = \lambda_k \qquad \text{for } k = 1, 2, \dots, p.$$

Geometrical interpretation



In general

Projection of the i-th sample $\mathbf{x}_i$ in the direction of α_1 is called ***score*** z_{i1}

Coefficients of the eigenvector are called loadings (are the weights multiplying the original variable to compute the PC)

If the variables in X have very different order of magnitude, we can work with the standardized variables, i.e., compute PCA on the correlation matrix. Results will be different (i.e, PCA on S is different than PCA on R)

In general

- The singular value decomposition (SVD) of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{\Phi}^T$$

where the ith column of $\mathbf{\Phi}$ is the ith principal component ϕ_i, and the ith column of $\mathbf{U}\mathbf{\Sigma}$ is the ith vector of scores $(z_{1i}, \ldots, z_{ni})$.

- The eigendecomposition of $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \mathbf{\Phi}\mathbf{\Sigma}^2\mathbf{\Phi}^T$$

http://web.stanford.edu/class/stats202/ - Data Mining and Analysis

How many PCs?

We can think of the top PCs as directions in space in which the data vary the most.

The i-th score vector $(z_{1i}, \ldots, z_{ni})$ can be interpreted as a new variable.

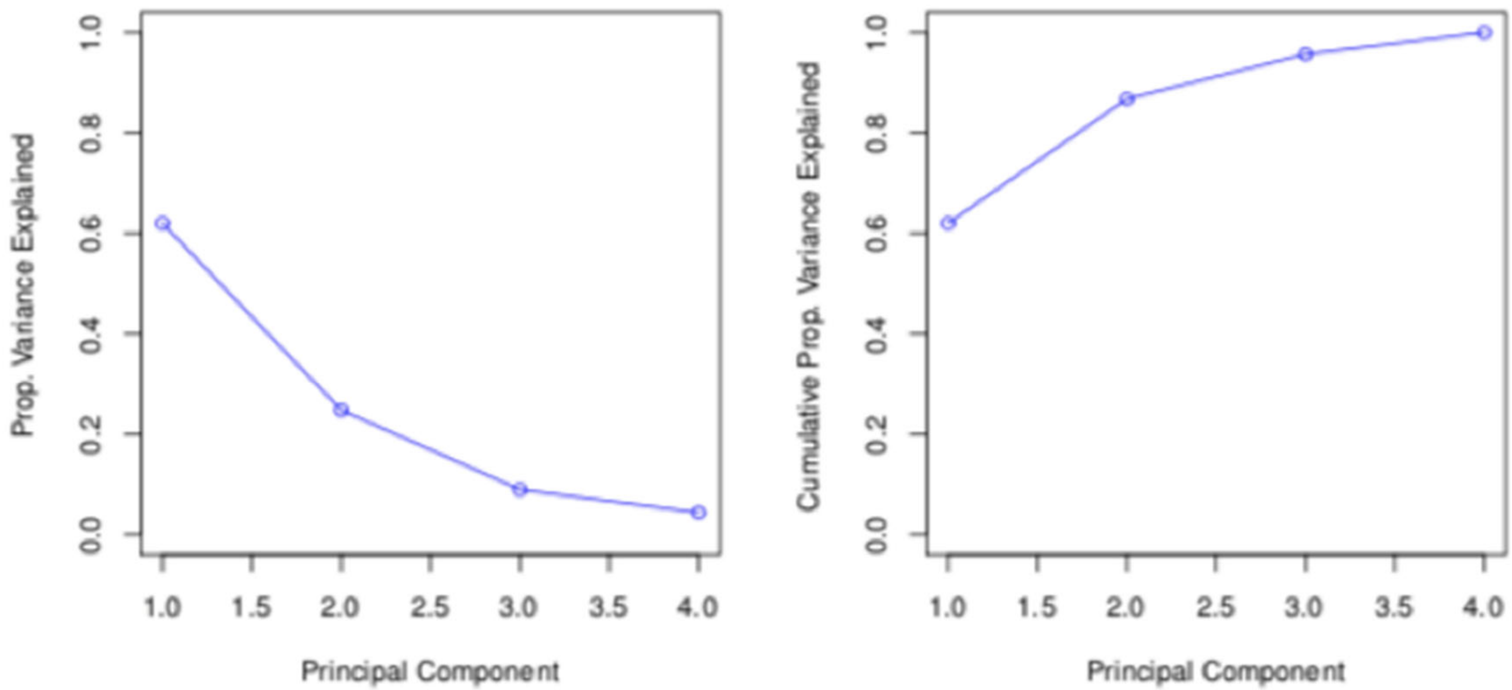
The variance of this variable decreases as we take i from 1 to p.

However, the total variance of the score vectors is the same as the total variance of the original variables:

$$\sum_{i=1}^{p} \frac{1}{n} \sum_{j=1}^{n} z_{ji}^2 = \sum_{k=1}^{p} \mathrm{Var}(x_k).$$

We can consider just the

Select the number of PCs



Scree plot

Principal Components

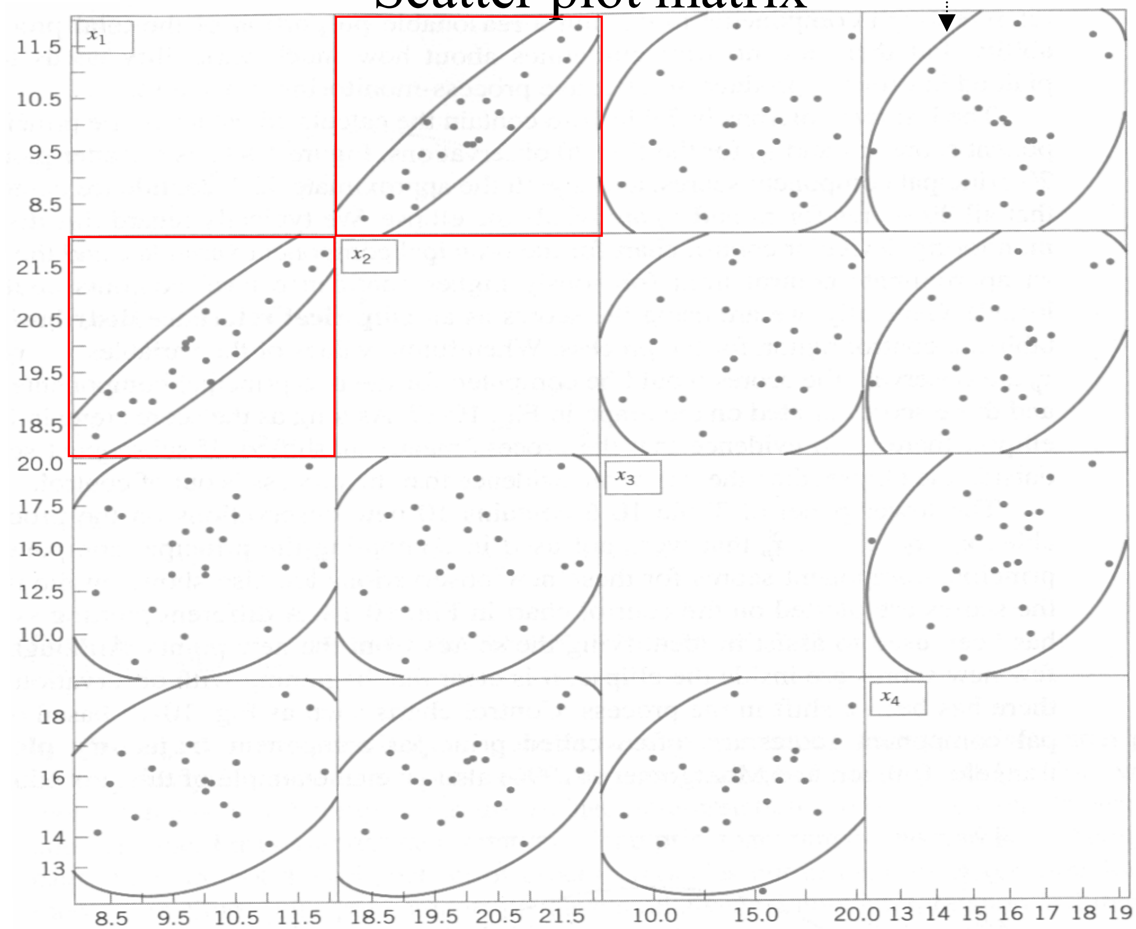
Example Montgomery: (chemical process)

	Original Data			
Observation	x_1	x_2	x_3	x_4
1	10	20.7	13.6	15.5
2	10.5	19.9	18.1	14.8
3	9.7	20	16.1	16.5
4	9.8	20.2	19.1	17.1
5	11.7	21.5	19.8	18.3
6	11	20.9	10.3	13.8
7	8.7	18.8	16.9	16.8
8	9.5	19.3	15.3	12.2
9	10.1	19.4	16.2	15.8
10	9.5	19.6	13.6	14.5
11	10.5	20.3	17	16.5
12	9.2	19	11.5	16.3
13	11.3	21.6	14	18.7
14	10	19.8	14	15.9
15	8.5	19.2	17.4	15.8
16	9.7	20.1	10	16.6
17	8.3	18.4	12.5	14.2
18	11.9	21.8	14.1	16.2
19	10.3	20.5	15.6	15.1
20	8.9	19	8.5	14.7

$$\mathbf{\Sigma} = \begin{bmatrix} 1.0000 & 0.9302 & 0.2060 & 0.3595 \\ 0.9302 & 1.0000 & 0.1669 & 0.4502 \\ 0.2060 & 0.1669 & 1.0000 & 0.3439 \\ 0.3595 & 0.4502 & 0.3439 & 1.0000 \end{bmatrix}$$

95% confidence region

Scatter plot matrix



(actually, it is the sample correlation matrix)

Latent Structure methods

Example:

"Principal Components Analysis" command (Minitab)

Table 10-7 PCA for the First 20 Observations on x_1, x_2, x_3, and x_4 from Table 10-6

Eigenvalues:	2.3181	1.0118	0.6088	0.0613
Percent:	57.9516	25.2951	15.2206	1.5328
Cumulative Percent:	57.9516	83.2466	98.4672	100.0000
Eigenvectors				
x_1	0.59410	−0.33393	0.25699	0.68519
x_2	0.60704	−0.32960	0.08341	−0.71826
x_3	0.28553	0.79369	0.53368	−0.06092
x_4	0.44386	0.38717	−0.80137	0.10440

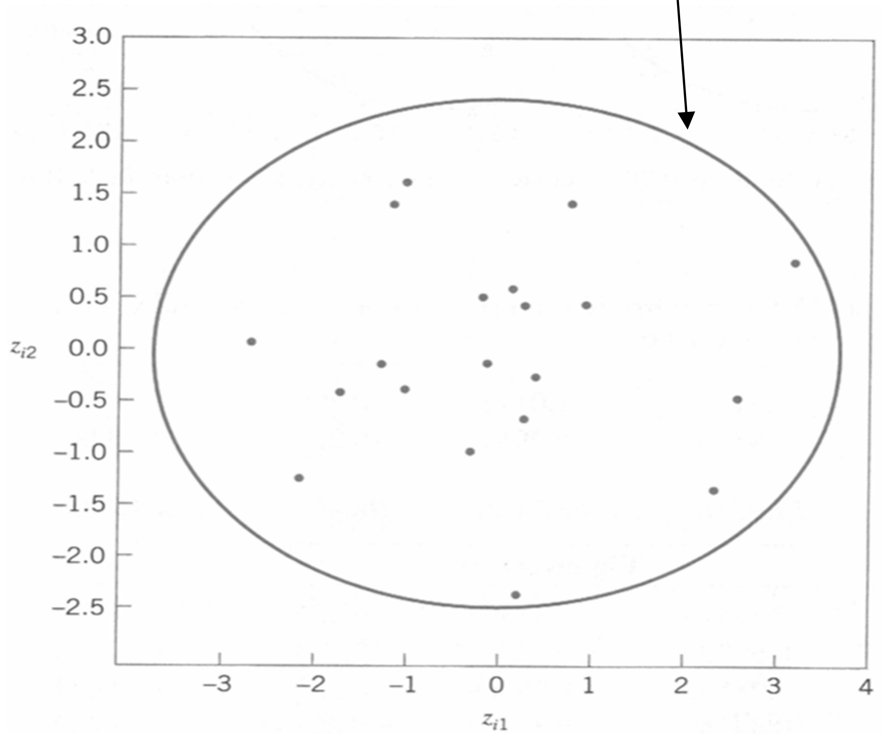
Principal components scores for first 20 observations:

z_1	z_2
0.291681	−0.6034
0.294281	0.491533
0.197337	0.640937
0.839022	1.469579
3.204876	0.879172
0.203271	−2.29514
−0.99211	1.670464
−1.70241	−0.36089
−0.14246	0.560808
−0.99498	−0.31493
0.944697	0.504711
−1.2195	−0.09129
2.608666	−0.42176
−0.12378	−0.08767
−1.10423	1.472593
−0.27825	−0.94763
−2.65608	0.135288
2.36528	−1.30494
0.411311	−0.21893
−2.14662	−1.17849

Latent Structure methods

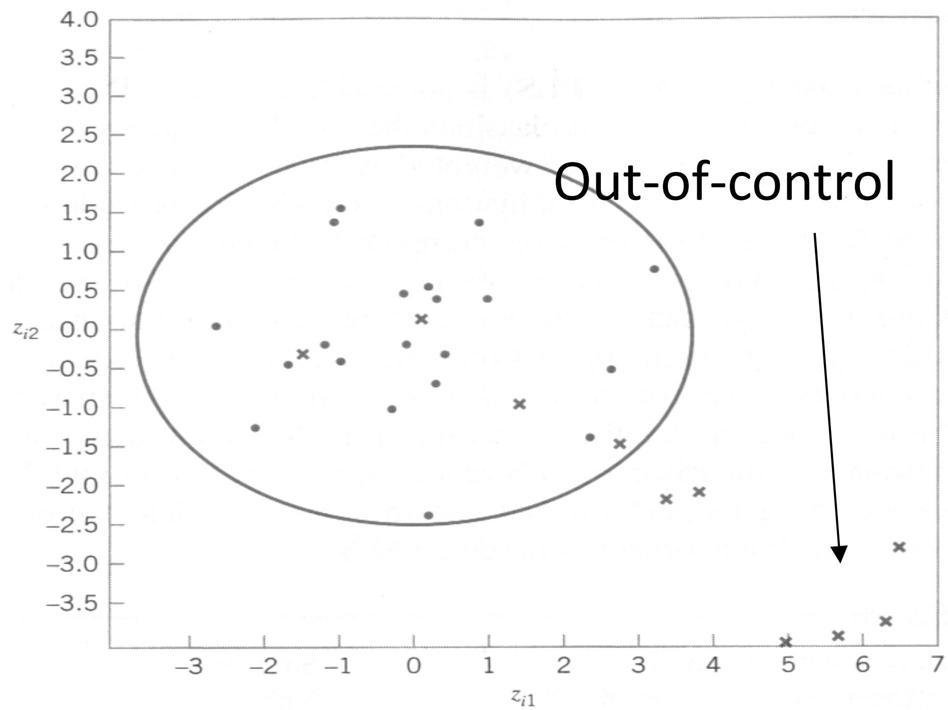
z_1	z_2
0.291681	− 0.6034
0.294281	0.491533
0.197337	0.640937
0.839022	1.469579
3.204876	0.879172
0.203271	− 2.29514
− 0.99211	1.670464
− 1.70241	− 0.36089
− 0.14246	0.560808
− 0.99498	− 0.31493
0.944697	0.504711
− 1.2195	− 0.09129
2.608666	− 0.42176
− 0.12378	− 0.08767
− 1.10423	1.472593
− 0.27825	− 0.94763
− 2.65608	0.135288
2.36528	− 1.30494
0.411311	− 0.21893
− 2.14662	− 1.17849

Scatter plot z_1- z_2 (with 95% confidence region: used as control region)



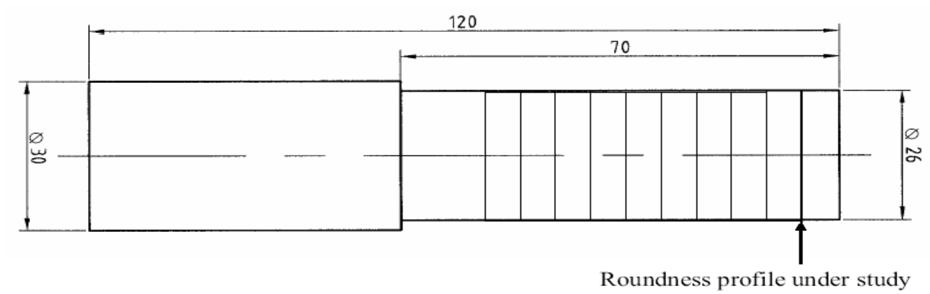
New Data

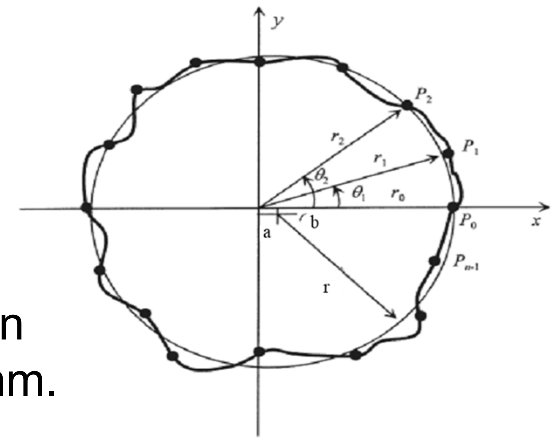
Observation	x_1	x_2	x_3	x_4	z_1	z_2
21	9.9	20	15.4	15.9	0.074196	0.239359
22	8.7	19	9.9	16.8	− 1.51756	− 0.21121
23	11.5	21.8	19.3	12.1	1.408476	− 0.87591
24	15.9	24.6	14.7	15.3	6.298001	− 3.67398
25	12.6	23.9	17.1	14.2	3.802025	− 1.99584
26	14.9	25	16.3	16.6	6.490673	− 2.73143
27	9.9	23.7	11.9	18.1	2.738829	− 1.37617
28	12.8	26.3	13.5	13.7	4.958747	− 3.94851
29	13.1	26.1	10.9	16.8	5.678092	− 3.85838
30	9.8	25.8	14.8	15	3.369657	− 2.10878



Example - PCA for profile modeling

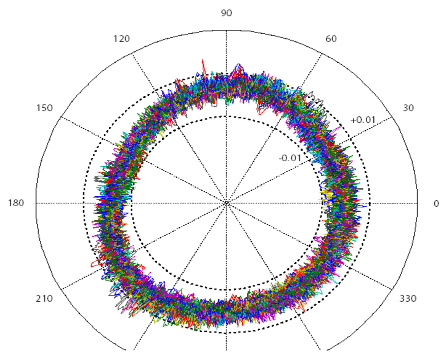
Colosimo Pacella "On the use of principal component analysis to identify systematic patterns in roundness profiles" 2007 - Quality and reliability eng. international

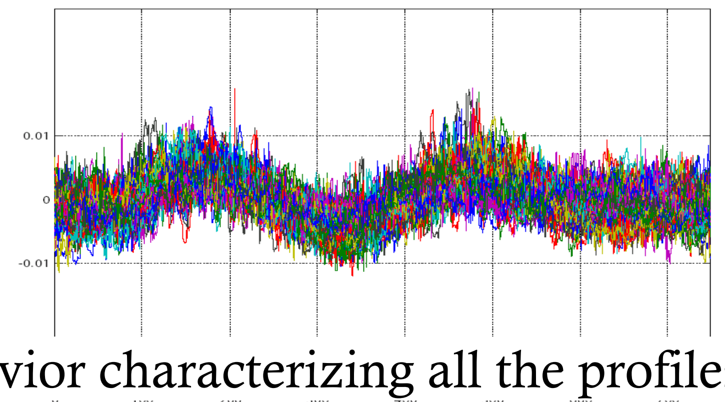




J=100 cast C20 carbon steel cylinders (supplied in Ø30 mm rolled bars) machined to nominal Ø26 mm.

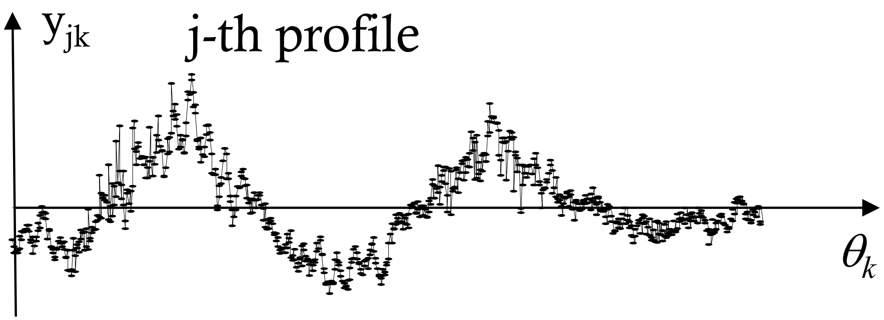
Each profile was sampled (t=1,…,748) by a CMM.

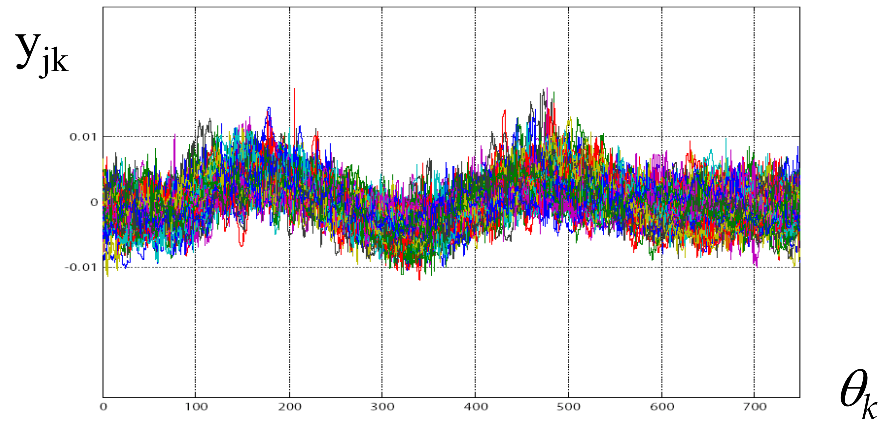




Manufacturing "signature": the systematic behavior characterizing all the profiles

Principal component analysis





$$\mathbf{y}_j^T = \begin{bmatrix} y_{j1} & y_{j2} & \ldots\ldots\ldots\ldots & y_{jp} \end{bmatrix}$$

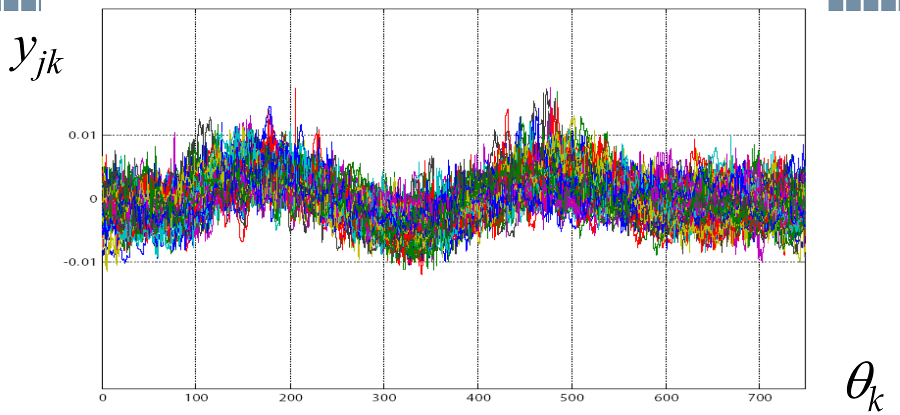
Main idea: reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set.

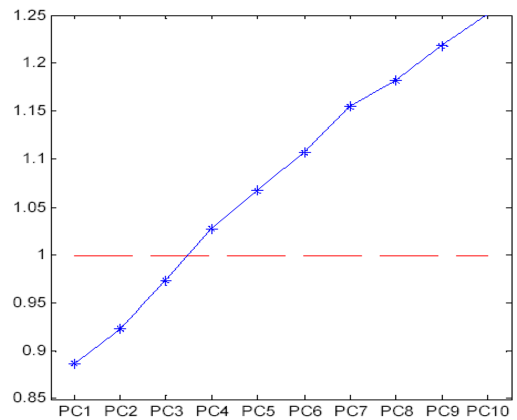
new set of variables (PCs) which are uncorrelated and which are ordered so the that the first few retain most of the variation present in all the original data. (Jolliffe, 2002)

$i = 1, \ldots, p;$ variable

$j = 1, \ldots, n$ sample

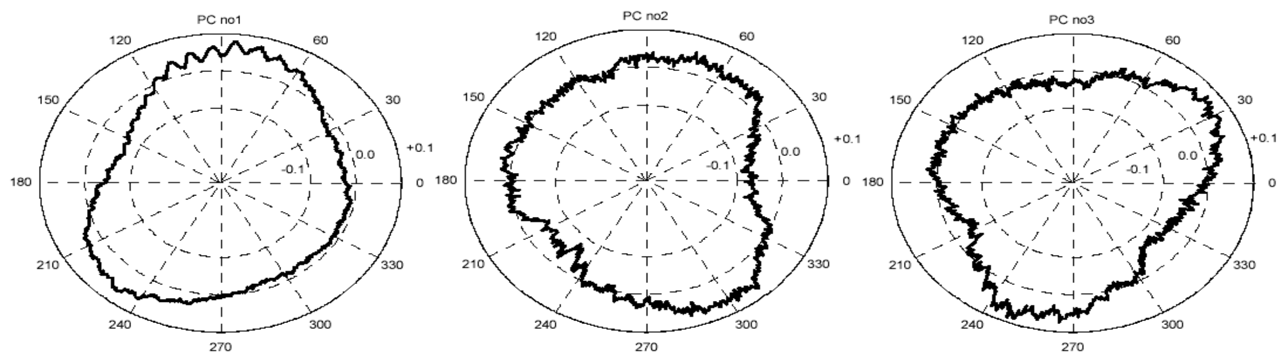
Real case and PCA-based control charting





Wold's R statistic versus the number of PCs.

Loadings of the first three PCs



Example – Hand-writing

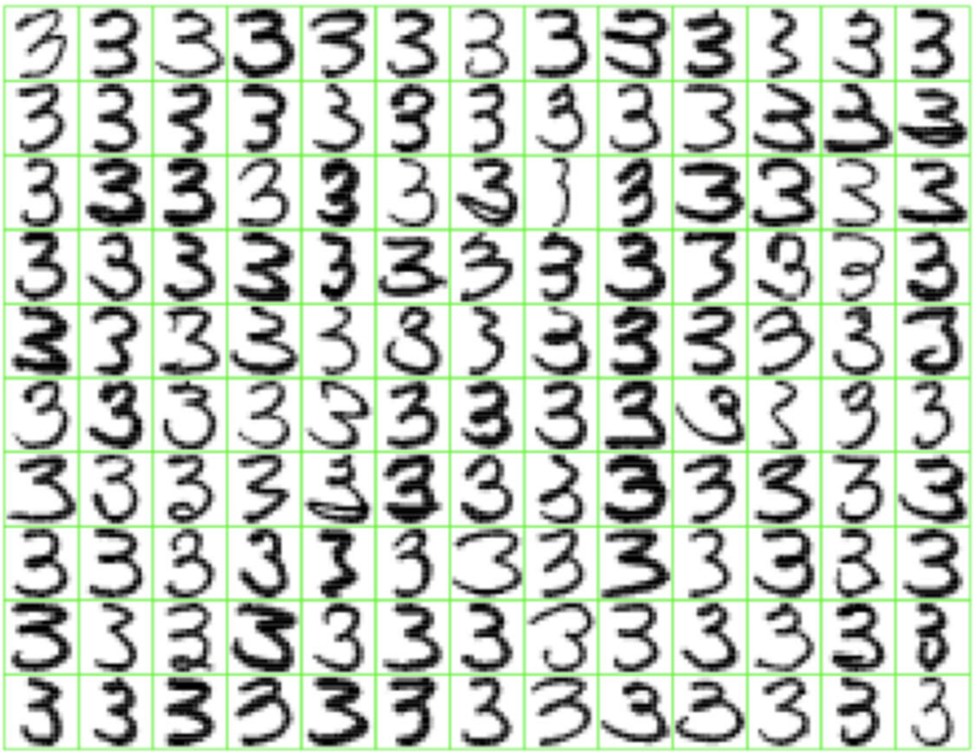


FIGURE 14.22. *A sample of 130 handwritten 3's shows a variety of writing styles.*

Principal components are a useful tool for dimension reduction and com pression.

In Hastie Tibshirani Friedman (HTF) this feature is shown on the handwritten digits data.

A sample of 130 handwritten 3's, each a digitized 16 × 16 grayscale image, from a total of 658 such 3's.

Considerable variation in writing styles, character thickness and orientation. We consider these images as points xi in $\mathbf{R}^{256}$, and compute their principal components via the SVD.

Hastie Tibshirani

We see that the v_1 (horizontal movement) mainly accounts for the lengthening of the lower tail of the three, while v_2 (vertical movement) accounts for character thickness.

$$\hat{f}(\lambda) = \bar{x} + \lambda_1 v_1 + \lambda_2 v_2$$

$=$  $+ \lambda_1 \cdot$  $+ \lambda_2 \cdot$  .

PCA for image data

http://people.ciirc.cvut.cz/~hlavac/TeachPresEn/11ImageProc/15PCA.pdf

- It took a while to realize (Turk, Pentland, 1991), but yes.
- Let us consider a 321×261 image.

- The image is considered as a very long 1D vector by concatenating image pixels column by column (or alternatively row by row), i.e. $321 \times 261 = 83781$.
- The huge number 83781 is the dimensionality of our vector space.
- The intensity variation is assumed in each pixel of the image.

What if we have 32 instances of images?

Fewer observations than unknowns, and what?

- We have only 32 observations and 83781 unknowns in our example!
- The induced system of linear equations is not over-constrained but under-constrained.
- PCA is still applicable.
- The number of principle components is less than or equal to the number of observations available (32 in our particular case). This is because the (square) covariance matrix has a size corresponding to the number of observations.
- The eigen-vectors we derive are called eigen-images, after rearranging back from the 1D vector to a rectangular image.
- Let us perform the dimensionality reduction from 32 to 4 in our example.

Approximation by 4 principal components only

- Reconstruction of the image from four basis vectors $\mathbf{b}_i$, $i = 1, \ldots, 4$ which can be displayed as images.
- The linear combination was computed as $q_1\mathbf{b}_1 + q_2\mathbf{b}_2 + q_3\mathbf{b}_3 + q_4\mathbf{b}_4 = 0.078\,\mathbf{b}_1 + 0.062\,\mathbf{b}_2 - 0.182\,\mathbf{b}_3 + 0.179\,\mathbf{b}_4$.

Reconstruction fidelity, 4 components

Reconstruction fidelity, original

PCA drawbacks, the images case

- By rearranging pixels column by column to a 1D vector, relations of a given pixel to pixels in neighboring rows are not taken into account.
- Another disadvantage is in the global nature of the representation; small change or error in the input images influences the whole eigen-representation. However, this property is inherent in all linear integral transforms.

Add on – reconstruct original information via PCA

$$\mathbf{X}=\begin{bmatrix} x_{11} & \cdots & x_{1i} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{j1} & \cdots & x_{ji} & \cdots & x_{jp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{ni} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_j^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \{x_{ji}\}$$

$$\mathbf{z}_j = \mathbf{U}^T(\mathbf{x}_j - \bar{\mathbf{x}}) = \mathbf{U}^T\tilde{\mathbf{x}}_j = \begin{bmatrix} z_{j1} & \cdots & z_{jk} & \cdots & z_{jp} \end{bmatrix}^T$$

Add on

$$\mathbf{U}^T\mathbf{S}_1\mathbf{U}=\mathbf{L}$$

Variance matrix

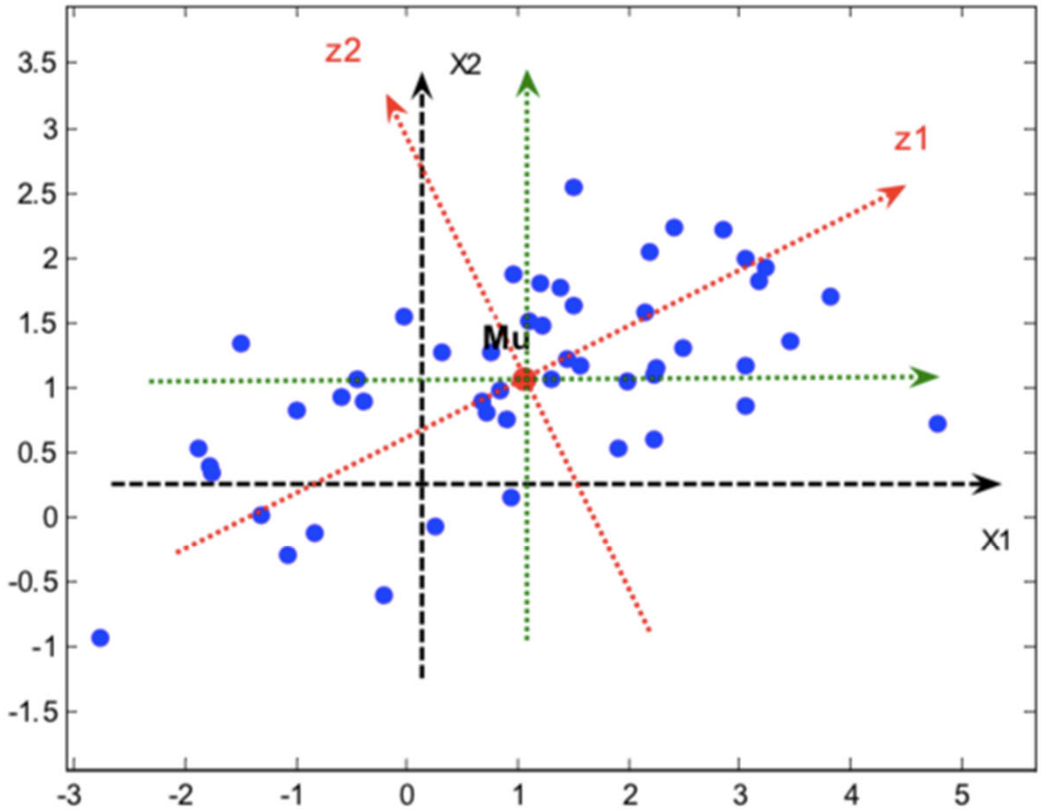
$$\mathbf{U}=\begin{bmatrix} u_{11} & \cdots & u_{1k} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{i1} & \vdots & u_{ik} & \cdots & u_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pk} & \cdots & u_{pp} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \cdots & \mathbf{u}_p \end{bmatrix}$$

Columns are the eigenvectors

$$\mathbf{L}=\begin{bmatrix} l_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \vdots & l_k & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & l_p \end{bmatrix}$$

eigenvalues

U is orthogonal $\quad \mathbf{U}^T=\mathbf{U}^{-1}$


z2
X2
z1
Mu
X1
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-3
-2
-1
0
1
2
3
4
5

Going back to the original system

$$\mathbf{x}_j = \begin{bmatrix} x_{j1} \\ \vdots \\ x_{jp} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} + \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & \ddots & \vdots \\ u_{p1} & \cdots & u_{pp} \end{bmatrix} \begin{bmatrix} z_{j1} \\ \vdots \\ z_{jp} \end{bmatrix} =$$

$$= \bar{\mathbf{x}} + z_{j1}\mathbf{u}_1 + z_{j2}\mathbf{u}_2 + \ldots + z_{jp}\mathbf{u}_p$$

$$\hat{\mathbf{x}}_j(K) = \begin{bmatrix} \hat{x}_{j1} & \cdots & \hat{x}_{jp} \end{bmatrix}^T = \bar{\mathbf{x}} + z_{j1}\mathbf{u}_1 + z_{j2}\mathbf{u}_2 + \ldots + z_{jK}\mathbf{u}_K$$