

Principal component analysis

David J. Hessen

Utrecht University

Academic year 2021-2022

Introduction

The $N \times p$ data **matrix**

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Np} \end{bmatrix}$$

Example 1: 9×2 data matrix

$$\mathbf{X} = \begin{bmatrix} 7.73 & 11.86 \\ 7.73 & 19.19 \\ 1.91 & 4.53 \\ 4.82 & 11.86 \\ 10.65 & 19.19 \\ 10.65 & 26.52 \\ 16.48 & 33.85 \\ 13.56 & 19.19 \\ 16.48 & 33.85 \end{bmatrix}$$

Introduction

The **transpose** of the data matrix

$$\mathbf{X}^T = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{N1} \\ x_{12} & x_{22} & \dots & x_{N2} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \dots & x_{Np} \end{bmatrix}$$

Example 1 (continued)

$$\mathbf{X}^T = \begin{bmatrix} 7.73 & 7.73 & 1.91 & 4.82 & 10.65 & 10.65 & 16.48 & 13.56 & 16.48 \\ 11.86 & 19.19 & 4.53 & 11.86 & 19.19 & 26.52 & 33.85 & 19.19 & 33.85 \end{bmatrix}$$

Introduction

The data **vector** of feature j is

$$\mathbf{x}_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Nj} \end{bmatrix}$$

The data **vector** of case i is

$$x_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip}]^T$$

So the data matrix equals

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_p] = [x_1 \ x_2 \ \dots \ x_N]^T$$

Introduction

The mean of feature j is

$$\bar{x}_j = \sum_{i=1}^N x_{ij}/N$$

The mean vector is

$$\bar{\mathbf{x}} = [\bar{x}_1 \ \dots \ \bar{x}_p]^T$$

Example 1 (continued)

$$\bar{\mathbf{x}} = [10 \ 20]^T$$

Introduction

The **centered** data vector of case i is

$$\tilde{x}_i = [\tilde{x}_{i1} \dots \tilde{x}_{ip}]^T = x_i - \bar{x} = [x_{i1} - \bar{x}_1 \dots x_{ip} - \bar{x}_p]^T$$

The centered data matrix is

$$\tilde{\mathbf{X}} = [\tilde{x}_1 \tilde{x}_2 \dots \tilde{x}_N]^T = [\tilde{\mathbf{X}}_1 \tilde{\mathbf{X}}_2 \dots \tilde{\mathbf{X}}_p]$$

Example 1 (continued)

$$\tilde{\mathbf{X}} = \begin{bmatrix} 7.73 - 10 & 11.86 - 20 \\ 7.73 - 10 & 19.19 - 20 \\ 1.91 - 10 & 4.53 - 20 \\ 4.82 - 10 & 11.86 - 20 \\ 10.65 - 10 & 19.19 - 20 \\ 10.65 - 10 & 26.52 - 20 \\ 16.48 - 10 & 33.85 - 20 \\ 13.56 - 10 & 19.19 - 20 \\ 16.48 - 10 & 33.85 - 20 \end{bmatrix} = \begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix}$$

Introduction

The sample covariance between features j and k is

$$s_{jk} = \tilde{\mathbf{x}}_j^T \tilde{\mathbf{x}}_k / N = \sum_{i=1}^N \tilde{x}_{ij} \tilde{x}_{ik} / N = (\tilde{x}_{1j} \tilde{x}_{1k} + \dots + \tilde{x}_{Nj} \tilde{x}_{Nk}) / N$$

where $\tilde{\mathbf{x}}_j^T \tilde{\mathbf{x}}_k$ is called the **dot product** of vectors $\tilde{\mathbf{x}}_j$ and $\tilde{\mathbf{x}}_k$

The sample variance of feature j is $s_{jj} = \tilde{\mathbf{x}}_j^T \tilde{\mathbf{x}}_j / N = s_j^2$

The sample covariance matrix is the **symmetric** matrix

$$\mathbf{S} = \begin{bmatrix} s_1^2 & & & \\ s_{21} & s_2^2 & & \\ \vdots & \vdots & \ddots & \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{bmatrix} = \mathbf{S}^T$$

Introduction

The sample covariance matrix equals

$$\mathbf{S} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} / N = \frac{1}{N} \begin{bmatrix} \tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_1 & & & \\ \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2^T \tilde{\mathbf{x}}_2 & & \\ \vdots & \vdots & \ddots & \\ \tilde{\mathbf{x}}_p^T \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_p^T \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_p^T \tilde{\mathbf{x}}_p \end{bmatrix}$$

The **total variance** is defined as the **trace** of \mathbf{S} given by

$$\text{tr}(\mathbf{S}) = \sum_{j=1}^p s_j^2 = s_1^2 + s_2^2 + \dots + s_p^2 = \frac{1}{N} \sum_{j=1}^p \tilde{\mathbf{x}}_j^T \tilde{\mathbf{x}}_j$$

Research question: Can most of the total variance be explained by a smaller than p number of dimensions?

Singular value decomposition

The matrix of centered data can be decomposed as

$$\tilde{\mathbf{X}} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

where

- ▶ $\mathbf{U} = [\mathbf{u}_1 \dots \mathbf{u}_p]$ is an $N \times p$ semi-orthogonal matrix whose columns are called the **left singular vectors**
- ▶ $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_p]$ is an $p \times p$ orthogonal matrix whose columns are called the **right singular vectors**
- ▶ \mathbf{D} is a $p \times p$ diagonal matrix with diagonal elements $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$ known as the **singular values**

Singular value decomposition

\mathbf{U} is a **semi-orthogonal** matrix, that is,

$$\mathbf{U}^T \mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \dots & \mathbf{u}_1^T \mathbf{u}_p \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \dots & \mathbf{u}_2^T \mathbf{u}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_p^T \mathbf{u}_1 & \mathbf{u}_p^T \mathbf{u}_2 & \dots & \mathbf{u}_p^T \mathbf{u}_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}$$

where \mathbf{I} is called the **identity** matrix

\mathbf{V} is an **orthogonal** matrix, that is, $\mathbf{V}^T \mathbf{V} = \mathbf{I} = \mathbf{V} \mathbf{V}^T$

Singular value decomposition

\mathbf{D} is a **diagonal** matrix, that is,

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_p \end{bmatrix}$$

where $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$

Singular value decomposition

An example of a 3×3 permutation matrix P

Let

$$\mathbf{D}_0 = \begin{bmatrix} d_2 & 0 & 0 \\ 0 & d_3 & 0 \\ 0 & 0 & d_1 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

where \mathbf{P} is an orthogonal **permutation** matrix, then

$$\mathbf{P}\mathbf{D}_0 = \begin{bmatrix} 0 & 0 & d_1 \\ d_2 & 0 & 0 \\ 0 & d_3 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{P}\mathbf{D}_0\mathbf{P}^T = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} = \mathbf{D}$$

$$\text{and } \tilde{\mathbf{X}} = \mathbf{U}_0\mathbf{D}_0\mathbf{V}_0^T = [\mathbf{u}_2 \ \mathbf{u}_3 \ \mathbf{u}_1]\mathbf{D}_0[\mathbf{v}_2 \ \mathbf{v}_3 \ \mathbf{v}_1]^T = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

$$\text{where } \mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3] = \mathbf{U}_0\mathbf{P}^T \quad \text{and} \quad \mathbf{V}^T = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]^T = \mathbf{P}\mathbf{V}_0^T$$

Singular value decomposition

Example 1 (continued)

$$\tilde{\mathbf{X}} = \begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.27 & 0.29 \\ -0.05 & -0.34 \\ -0.56 & -0.13 \\ -0.31 & -0.24 \\ -0.01 & 0.19 \\ 0.20 & -0.44 \\ 0.49 & -0.02 \\ 0.03 & 0.71 \\ 0.49 & -0.02 \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} 31.22 & 0.00 \\ 0.00 & 5.03 \end{bmatrix}}_{\mathbf{D}} \underbrace{\begin{bmatrix} 0.43 & 0.90 \\ 0.90 & -0.43 \end{bmatrix}}_{\mathbf{V}^T}$$

Principal components

The elements of the columns of the $N \times p$ matrix

$$\tilde{\mathbf{X}}\mathbf{V} = \mathbf{U}\mathbf{D}$$

are the principal component scores

This matrix of principal component data is denoted by

$$\hat{\mathbf{\Lambda}} = \begin{bmatrix} \hat{\lambda}_{11} & \dots & \hat{\lambda}_{1p} \\ \vdots & & \vdots \\ \hat{\lambda}_{N1} & \dots & \hat{\lambda}_{Np} \end{bmatrix}$$

The value of case i on principal component j is

$$\hat{\lambda}_{ij} = v_{1j}\tilde{x}_{i1} + v_{2j}\tilde{x}_{i2} + \dots + v_{pj}\tilde{x}_{ip} = \mathbf{v}_j^T \tilde{\mathbf{x}}_i = u_{ij}d_j$$

Principal components

Example 1 (continued)

$$\hat{\Lambda} = \underbrace{\begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix}}_{\tilde{X}V} \begin{bmatrix} 0.43 & 0.90 \\ 0.90 & -0.43 \end{bmatrix} = \underbrace{\begin{bmatrix} -0.27 & 0.29 \\ -0.05 & -0.34 \\ -0.56 & -0.13 \\ -0.31 & -0.24 \\ -0.01 & 0.19 \\ 0.20 & -0.44 \\ 0.49 & -0.02 \\ 0.03 & 0.71 \\ 0.49 & -0.02 \end{bmatrix}}_{UD} \begin{bmatrix} 31.22 & 0.00 \\ 0.00 & 5.03 \end{bmatrix} = \begin{bmatrix} -8.33 & 1.45 \\ -1.71 & -1.70 \\ -17.45 & -0.67 \\ -9.58 & -1.19 \\ -0.46 & 0.93 \\ 6.16 & -2.21 \\ 15.28 & -0.09 \\ 0.79 & 3.57 \\ 15.28 & -0.09 \end{bmatrix}$$

Principal components

Since the mean of the values on the **first** principal component is zero, the variance equals

$$\frac{1}{N} \sum_{i=1}^N \hat{\lambda}_{i1}^2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{v}_1^T \tilde{x}_i)^2 = \frac{d_1^2}{N}$$

The elements of \mathbf{v}_1 are chosen such that this variance is maximum, subject to the constraint that $\mathbf{v}_1^T \mathbf{v}_1 = 1$

Principal components

Since the mean of the values on the **second** principal component is zero, the variance equals

$$\frac{1}{N} \sum_{i=1}^N \hat{\lambda}_{i2}^2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{v}_2^T \tilde{x}_i)^2 = \frac{d_2^2}{N}$$

The elements of \mathbf{v}_2 are chosen such that this variance is maximum, subject to the constraints that $\mathbf{v}_2^T \mathbf{v}_2 = 1$ and the values on this principal component are uncorrelated with the values on the first principal component

Principal components

Since the mean of the values on the **third** principal component is zero, the variance equals

$$\frac{1}{N} \sum_{i=1}^N \hat{\lambda}_{i3}^2 = \frac{1}{N} \sum_{i=1}^N (\mathbf{v}_3^T \tilde{x}_i)^2 = \frac{d_3^2}{N}$$

The elements of \mathbf{v}_3 are chosen such that this variance is maximum, subject to the constraints that $\mathbf{v}_3^T \mathbf{v}_3 = 1$ and the values on this principal component are uncorrelated with the values on the first and second principal components

And so on up to the p th principal component

Principal components

Example 1 (continued)

Two **centered** features

i	\tilde{x}_{i1}	\tilde{x}_{i2}
1	-2.27	-8.14
2	-2.27	-0.81
3	-8.09	-15.47
4	-5.18	-8.14
case 5	0.65	-0.81
6	0.65	6.52
7	6.48	13.85
8	3.56	-0.81
9	6.48	13.85

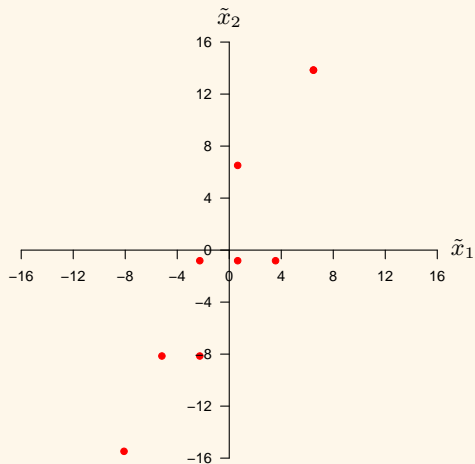
In the case of 2 features, 2 principal components are constructed

$$\hat{\lambda}_{i1} = \mathbf{v}_1^T \tilde{\mathbf{x}}_i$$

$$\hat{\lambda}_{i2} = \mathbf{v}_2^T \tilde{\mathbf{x}}_i$$

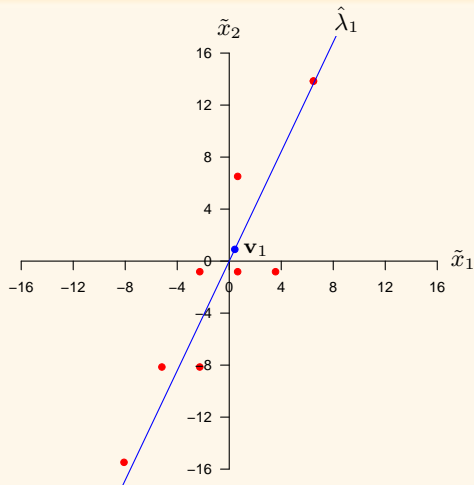
Principal components

Example 1 (continued)



Principal components

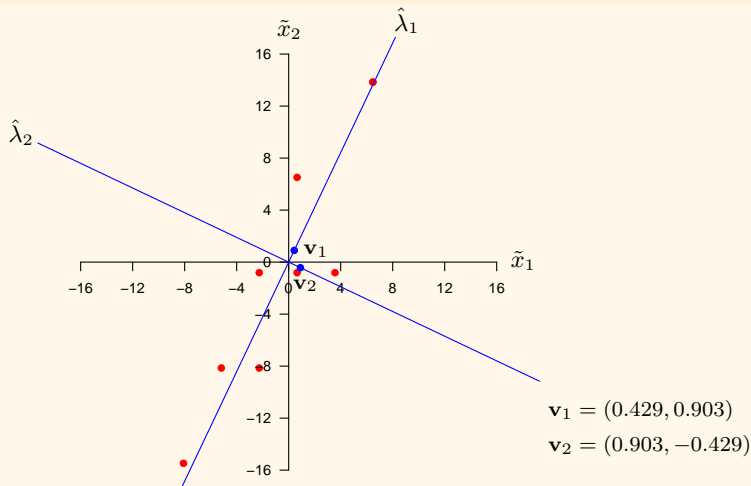
Example 1 (continued)



$$\mathbf{v}_1 = (0.429, 0.903)$$

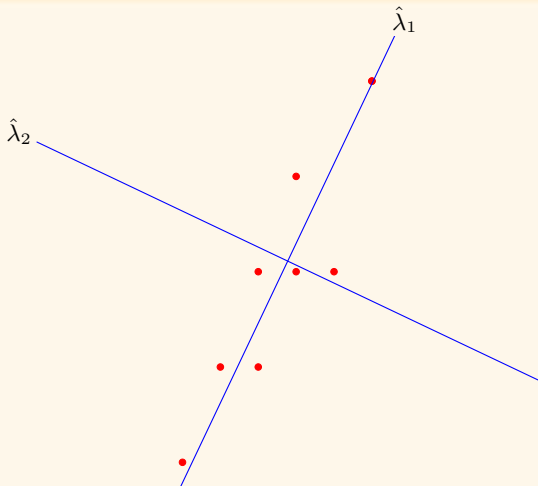
Principal components

Example 1 (continued)



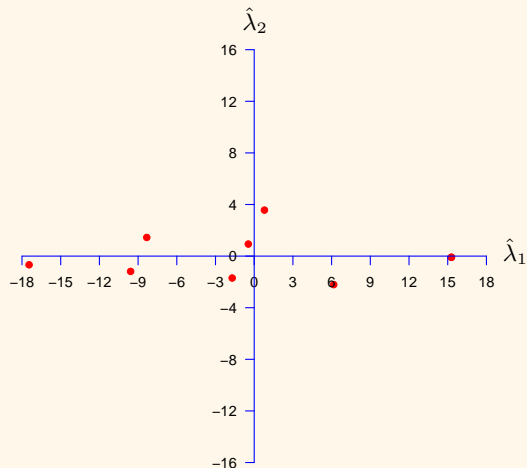
Principal components

Example 1 (continued)



Principal components

Example 1 (continued)



Principal components

Example 1 (continued)

The two principal components are

$$\hat{\lambda}_{i1} = 0.429\tilde{x}_{i1} + 0.903\tilde{x}_{i2}$$

$$\hat{\lambda}_{i2} = 0.903\tilde{x}_{i1} - 0.429\tilde{x}_{i2}$$

<i>i</i>	\tilde{x}_{i1}	\tilde{x}_{i2}	$\hat{\lambda}_{i1}$	$\hat{\lambda}_{i2}$
1	-2.27	-8.14	-8.33	1.45
2	-2.27	-0.81	-1.71	-1.70
3	-8.09	-15.47	-17.45	-0.67
4	-5.18	-8.14	-9.58	-1.19
case 5	0.65	-0.81	-0.46	0.93
6	0.65	6.52	6.16	-2.21
7	6.48	13.85	15.28	-0.09
8	3.56	-0.81	0.79	3.57
9	6.48	13.85	15.28	-0.09

Eigendecomposition

It follows that the sample covariance matrix equals

$$\mathbf{S} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}/N = (\mathbf{U}\mathbf{D}\mathbf{V}^T)^T \mathbf{U}\mathbf{D}\mathbf{V}^T/N = \mathbf{V}\mathbf{D}^2\mathbf{V}^T/N = \mathbf{V}\mathbf{\Delta}\mathbf{V}^T$$

where

- ▶ the columns of $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_p]$ are now called the **eigenvectors** of covariance matrix \mathbf{S}
- ▶ $\mathbf{\Delta} = \mathbf{D}^2/N$ is a $p \times p$ diagonal matrix with diagonal elements

$$\delta_1 = d_1^2/N \geq \delta_2 = d_2^2/N \geq \dots \geq \delta_p = d_p^2/N \geq 0$$

known as the **eigenvalues** of covariance matrix \mathbf{S} (the variances of the principal components)

Eigendecomposition

Example 1 (continued)

$$\mathbf{S} = \frac{1}{9} \begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix}^T \begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix} = \begin{bmatrix} 22.22 & 40.87 \\ 40.87 & 88.89 \end{bmatrix}$$
$$\mathbf{S} = \begin{bmatrix} 22.22 & 40.87 \\ 40.87 & 88.89 \end{bmatrix} = \begin{bmatrix} 0.43 & 0.90 \\ 0.90 & -0.43 \end{bmatrix} \begin{bmatrix} 108.30 & 0.00 \\ 0.00 & 2.81 \end{bmatrix} \begin{bmatrix} 0.43 & 0.90 \\ 0.90 & -0.43 \end{bmatrix}$$

Total variance explained

Can most of the total variance be explained by a smaller than p number of principal components?

Total variance

$$\sum_{j=1}^p s_j^2 = \text{tr}(\mathbf{S}) = \text{tr}(\mathbf{V}\mathbf{\Delta}\mathbf{V}^T) = \text{tr}(\mathbf{\Delta}\mathbf{V}^T\mathbf{V}) = \text{tr}(\mathbf{\Delta}) = \sum_{j=1}^p \delta_j$$

The percentage of total variance explained by the j th principal component is

$$\{\delta_j / \text{tr}(\mathbf{\Delta})\} \times 100\%$$

The **cumulative** percentage of total variance explained by the first q principal components is

$$\{(\delta_1 + \dots + \delta_q) / \text{tr}(\mathbf{\Delta})\} \times 100\%$$

Total variance explained

Example 1 (continued)

	i	\tilde{x}_{i1}	\tilde{x}_{i2}	$\hat{\lambda}_{i1}$	$\hat{\lambda}_{i2}$
	1	-2.27	-8.14	-8.33	1.45
	2	-2.27	-0.81	-1.71	-1.70
	3	-8.09	-15.47	-17.45	-0.67
	4	-5.18	-8.14	-9.58	-1.19
case	5	0.65	-0.81	-0.46	0.93
	6	0.65	6.52	6.16	-2.21
	7	6.48	13.85	15.28	-0.09
	8	3.56	-0.81	0.79	3.57
	9	6.48	13.85	15.28	-0.09

Eigenvalues $\delta_1 = 108.30$ and $\delta_2 = 2.81$

Total variance explained

The number of principal components to be extracted is equal to the number of principal components with a cumulative percentage of total variance explained at least as high as a prespecified percentage

Example 1 (continued)

Suppose it is desired to explain at least 80% of the total variance

The percentage of total variance explained by the first principal component is

$$\frac{108.30}{108.30 + 2.81} \times 100\% \approx 97\%$$

According to this criterion, one principal component should be extracted

Standardization

Let $S = \text{diag}\{s_1, s_2, \dots, s_p\}$

The **inverse** of S is $S^{-1} = \text{diag}\left\{\frac{1}{s_1}, \frac{1}{s_2}, \dots, \frac{1}{s_p}\right\}$ because $SS^{-1} = \mathbf{I}$

The **standardized** data matrix is

$$\mathbf{Z} = \tilde{\mathbf{X}}S^{-1}$$

The covariance matrix of the standardized features is the **correlation matrix**

$$\mathbf{R} = \mathbf{Z}^T\mathbf{Z}/N = (\tilde{\mathbf{X}}S^{-1})^T\tilde{\mathbf{X}}S^{-1}/N = S^{-1}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}/N)S^{-1} = S^{-1}\mathbf{S}S^{-1}$$

Standardization

Example 1 (continued)

The standardized data matrix

$$\mathbf{Z} = \begin{bmatrix} -2.27 & -8.14 \\ -2.27 & -0.81 \\ -8.09 & -15.47 \\ -5.18 & -8.14 \\ 0.65 & -0.81 \\ 0.65 & 6.52 \\ 6.48 & 13.85 \\ 3.56 & -0.81 \\ 6.48 & 13.85 \end{bmatrix} \begin{bmatrix} 4.71 & 0.00 \\ 0.00 & 9.43 \end{bmatrix}^{-1} = \begin{bmatrix} -0.48 & -0.86 \\ -0.48 & -0.09 \\ -1.72 & -1.64 \\ -1.10 & -0.86 \\ 0.14 & -0.09 \\ 0.14 & 0.69 \\ 1.37 & 1.47 \\ 0.76 & -0.09 \\ 1.37 & 1.47 \end{bmatrix}$$

The correlation matrix

$$\mathbf{R} = \mathbf{Z}^T \mathbf{Z} / 9 = \begin{bmatrix} 1.00 & 0.92 \\ 0.92 & 1.00 \end{bmatrix}$$

Singular value decomposition of \mathbf{Z}

Example 1 (continued)

$$\mathbf{Z} = \begin{bmatrix} -0.48 & -0.86 \\ -0.48 & -0.09 \\ -1.72 & -1.64 \\ -1.10 & -0.86 \\ 0.14 & -0.09 \\ 0.14 & 0.69 \\ 1.37 & 1.47 \\ 0.76 & -0.09 \\ 1.37 & 1.47 \end{bmatrix} = \begin{bmatrix} -0.23 & 0.32 \\ -0.10 & -0.33 \\ -0.57 & -0.06 \\ -0.33 & -0.20 \\ 0.01 & 0.19 \\ 0.14 & -0.46 \\ 0.48 & -0.08 \\ 0.11 & 0.70 \\ 0.48 & -0.08 \end{bmatrix} \begin{bmatrix} 4.16 & 0.00 \\ 0.00 & 0.85 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \\ 0.71 & -0.71 \end{bmatrix}$$

Singular value decomposition of \mathbf{Z}

Example 1 (continued)

Principal component scores

$$\tilde{\mathbf{A}} = \begin{bmatrix} -0.48 & -0.86 \\ -0.48 & -0.09 \\ -1.72 & -1.64 \\ -1.10 & -0.86 \\ 0.14 & -0.09 \\ 0.14 & 0.69 \\ 1.37 & 1.47 \\ 0.76 & -0.09 \\ 1.37 & 1.47 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \\ 0.71 & -0.71 \end{bmatrix} = \begin{bmatrix} -0.23 & 0.32 \\ -0.10 & -0.33 \\ -0.57 & -0.06 \\ -0.33 & -0.20 \\ 0.01 & 0.19 \\ 0.14 & -0.46 \\ 0.48 & -0.08 \\ 0.11 & 0.70 \\ 0.48 & -0.08 \end{bmatrix} \begin{bmatrix} 4.16 & 0.00 \\ 0.00 & 0.85 \end{bmatrix} = \begin{bmatrix} -0.95 & 0.27 \\ -0.40 & -0.28 \\ -2.37 & -0.05 \\ -1.39 & -0.17 \\ 0.04 & 0.16 \\ 0.59 & -0.39 \\ 2.01 & -0.07 \\ 0.47 & 0.60 \\ 2.01 & -0.07 \end{bmatrix}$$

Eigendecomposition of \mathbf{R}

Example 1 (continued)

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.92 \\ 0.92 & 1.00 \end{bmatrix} = \begin{bmatrix} 0.71 & 0.71 \\ 0.71 & -0.71 \end{bmatrix} \begin{bmatrix} 1.92 & 0.00 \\ 0.00 & 0.08 \end{bmatrix} \begin{bmatrix} 0.71 & 0.71 \\ 0.71 & -0.71 \end{bmatrix}$$

Total variance: $\text{tr}(\mathbf{R}) = p$

The percentage of total variance explained by the first principal component is

$$\frac{1.92}{1.92 + 0.08} \times 100\% \approx 96\%$$

Scree test

Example 2

6 standardized features and $n = 1000$

The first three are measurements of spacial ability

The last three are measurements of verbal ability

Correlation matrix

$$\begin{bmatrix} 1.00 & & & & & \\ 0.75 & 1.00 & & & & \\ 0.81 & 0.78 & 1.00 & & & \\ 0.18 & 0.25 & 0.25 & 1.00 & & \\ 0.08 & 0.15 & 0.15 & 0.90 & 1.00 & \\ 0.14 & 0.21 & 0.20 & 0.87 & 0.84 & 1.00 \end{bmatrix}$$

Scree test

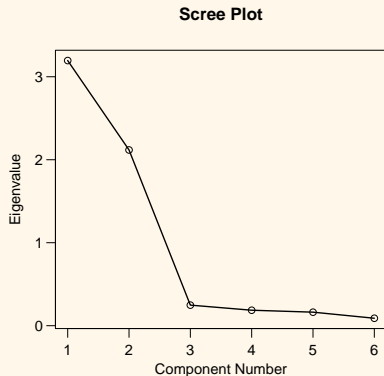
Example 2 (continued)

Component	Eigenvalue	PVE	CPVE
1	3.195	53.256	53.256
2	2.118	35.296	88.552
3	.248	4.130	92.682
4	.186	3.107	95.789
5	.163	2.715	98.504
6	.090	1.496	100.000

In a so-called scree plot, the eigenvalues of the principal components are plotted against the rank numbers of the principal components

Scree test

Example 2 (continued)



The number of principal components to be extracted is equal to the number of eigenvalues greater than the elbow in the scree plot

Interpretation

The **component matrix** might help in interpreting the extracted principal components → which features play a role?

The elements of the component matrix are

- ▶ the correlations between the standardized features and the extracted principal components
- ▶ the standardized regression coefficients from the regression of the standardized features on the extracted principal components

	Component		
	1	...	q
Standardized feature	r_{11}	...	r_{1q}
	2	...	r_{2q}
	\vdots		\vdots
	p	...	r_{pq}

Interpretation

Example 2 (continued)

	Component	
	1	2
Feature 1	.630	.678
Feature 2	.682	.609
Feature 3	.688	.633
Feature 4	.825	-.504
Feature 5	.755	-.593
Feature 6	.781	-.531

Principal component regression

Supervised learning

- ▶ a single response y
- ▶ m principal components of the features x_1, \dots, x_p as predictors

Multiple regression (interval response): $y = f(\hat{\lambda}_1, \dots, \hat{\lambda}_m) + \varepsilon$

Binary logistic regression: $\pi = \frac{\exp\{f(\hat{\lambda}_1, \dots, \hat{\lambda}_m)\}}{1 + \exp\{f(\hat{\lambda}_1, \dots, \hat{\lambda}_m)\}}$

The number of principal components m can be determined by cross-validation