Module 18 Multivariate Analysis for Genetic data
Session 02: Matrix decompositions

Jan Graffelman

jan.graffelman@upc.edu

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Barcelona, Spain

[2]Department of Biostatistics
University of Washington
Seattle, WA, USA

26th Summer Institute in Statistical Genetics (SISG 2021)

Introduction

- Introduction
- The spectral decomposition
- The singular value decomposition
- Numerical examples

Matrix decompositions

- Matrix decompositions play an important role in multivariate analysis.
- These decompositions allow the approximation of high-dimensional data sets in fewer dimensions.
- Matrix decompositions are the mathematical underpinning of multivariate graphics called biplots (next session).

Matrix decompositions

- We will discuss two decompositions
- The spectral decomposition (or eigenvalue-eigenvector decomposition or eigendecomposition)
- The singular value decomposition (SVD or Eckart-Young theorem)
- Many classical multivariate methods (PCA, CA, MDS, CCA, LDA) are based on these decompositions.

Correlation matrix of allele intensities

	SNP1_IA	SNP1_IB	SNP2_IA	SNP2_IB
SNP1_IA	1.00	-0.89	0.96	-0.94
SNP1_IB	-0.89	1.00	-0.89	0.96
SNP2_IA	0.96	-0.89	1.00	-0.92
SNP2_IB	-0.94	0.96	-0.92	1.00

The problem

- The correlation matrix **R** is 4 × 4, and of rank 4.
- Can we approximate **R** by a rank 2 matrix, say $\hat{\mathbf{R}}$
- Entries of $\hat{\mathbf{R}}$ must be as "close" as possible to **R**
- A rank 2 matrix can be represented exactly in a two-dimensional graph.

The Solution

$$\mathbf{R} = \begin{bmatrix} 1.00 & -0.89 & 0.96 & -0.94 \\ -0.89 & 1.00 & -0.89 & 0.96 \\ 0.96 & -0.89 & 1.00 & -0.92 \\ -0.94 & 0.96 & -0.92 & 1.00 \end{bmatrix}$$

$$\hat{\mathbf{R}} = \begin{bmatrix} 0.98 & -0.90 & 0.98 & -0.94 \\ -0.90 & 0.99 & -0.88 & 0.98 \\ 0.98 & -0.88 & 0.98 & -0.93 \\ -0.94 & 0.98 & -0.93 & 0.98 \end{bmatrix}$$

Least squares criterion

- In linear regression, we estimate the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ by minimizing $\sum e_i^2$, where $e_i = y_i - (b_0 + b_1 x_i)$.
- In this matrix approximation we minimize the errors in $\mathbf{E} = \mathbf{R} - \hat{\mathbf{R}}$.
- The least squares criterion amounts to $\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 = \mathrm{tr}(\mathbf{E}'\mathbf{E})$.
- The approximation $\hat{\mathbf{R}}$ can be obtained by extracting eigenvalues and eigenvectors of $\mathbf{R}$.

Eigenvalues & eigenvectors

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \qquad \mathbf{A}_{k\times k}$$

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$$

The characteristic equation

$$|\,\mathbf{A} - \lambda\mathbf{I}\,| = 0$$

- There are k roots, not necessarily all distinct.
- The roots are called eigenvalues or characteristic values
- If $\mathbf{A} = \mathbf{A}'$, all roots are real.
- Each root (eigenvalue) has an associated eigenvector or characteristic vector.
- $\mathbf{v}$ usually scaled (normalized) to unit length such that $\mathbf{v}'\mathbf{v} = 1$.

Spectral decomposition

$$\mathbf{A}_{k\times k} \text{ and } \mathbf{A} = \mathbf{A}'$$

$$\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i \mathbf{v}_i' = \lambda_1 \mathbf{v}_1 \mathbf{v}_1' + \lambda_2 \mathbf{v}_2 \mathbf{v}_2' + \cdots + \lambda_k \mathbf{v}_k \mathbf{v}_k'$$

$$\mathbf{A} = \mathbf{V}\mathbf{D}_\lambda \mathbf{V}'$$

$$\mathbf{V} = \left[\ \mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_k\ \right], \quad \mathbf{D}_\lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}, \quad \mathbf{V}'\mathbf{V} = \mathbf{I}.$$

- Eigenvalues usually ordered s.t. $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k$
- If **A** is not of full rank, there will be zero eigenvalues.
- $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{V}\mathbf{D}_\lambda\mathbf{V}') = \mathrm{tr}(\mathbf{V}'\mathbf{V}\mathbf{D}_\lambda) = \mathrm{tr}(\mathbf{I}\mathbf{D}_\lambda) = \mathrm{tr}(\mathbf{D}_\lambda) = \sum_{i=1}^{k} \lambda_i$.
- $\lambda_1 \mathbf{v}_1 \mathbf{v}_1' + \lambda_2 \mathbf{v}_2 \mathbf{v}_2'$ provides a rank 2 least squares approximation to a **A**.

Matrix powers

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}'$$

$$\mathbf{A}^2 = \mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{V}\mathbf{D}\mathbf{V}' = \mathbf{V}\mathbf{D}^2\mathbf{V}'$$

$$\mathbf{A}^k = \mathbf{V}\mathbf{D}^k\mathbf{V}'$$

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{V}'$$

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{V}\mathbf{D}^{\frac{1}{2}}\mathbf{V}'$$

$$\mathbf{A}^{-\frac{1}{2}} = \mathbf{V}\mathbf{D}^{-\frac{1}{2}}\mathbf{V}'$$

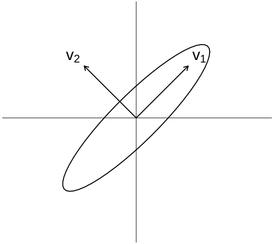
Spectral decomposition in R

```
> X <- read.table("http://www-eio.upc.es/~jan/SISG/Intensities.dat",header=TRUE)
> R <- cor(X[,1:4])
> R
           SNP1_IA    SNP1_IB    SNP2_IA    SNP2_IB
SNP1_IA  1.0000000 -0.8947940  0.9617704 -0.9367175
SNP1_IB -0.8947940  1.0000000 -0.8859773  0.9618440
SNP2_IA  0.9617704 -0.8859773  1.0000000 -0.9204350
SNP2_IB -0.9367175  0.9618440 -0.9204350  1.0000000
>
> out <- eigen(R)
> V <- out$vectors
> V
           [,1]       [,2]       [,3]       [,4]
[1,]  0.5016859 -0.4352583 -0.5749687  0.4777787
[2,] -0.4948632 -0.6497792 -0.3320110 -0.4718751
[3,]  0.4983229 -0.5334929  0.6404236 -0.2385733
[4,] -0.5050703 -0.3220598  0.3860534  0.7015299
> Dl <- diag(out$values)
> Dl
         [,1]      [,2]       [,3]       [,4]
[1,] 3.780985 0.0000000 0.00000000 0.00000000
[2,] 0.000000 0.1499301 0.00000000 0.00000000
[3,] 0.000000 0.0000000 0.04099305 0.00000000
[4,] 0.000000 0.0000000 0.00000000 0.02809177
>
```

Spectral decomposition in R

```
> t(V)%*%V
              [,1]          [,2]          [,3]         [,4]
[1,]  1.000000e+00 -1.665335e-16 -2.775558e-16  0.00000e+00
[2,] -1.665335e-16  1.000000e+00  5.551115e-17 -1.94289e-16
[3,] -2.775558e-16  5.551115e-17  1.000000e+00  0.00000e+00
[4,]  0.000000e+00 -1.942890e-16  0.000000e+00  1.00000e+00
>
> V2 <- V[,1:2]
> Dl2 <- Dl[1:2,1:2]
>
> Rhat <- V2%*%Dl2%*%t(V2)
>
> Rhat
           [,1]       [,2]       [,3]       [,4]
[1,]  0.9800356 -0.8962861  0.9800670 -0.9370340
[2,] -0.8962861  0.9892262 -0.8804235  0.9763975
[3,]  0.9800670 -0.8804235  0.9815881 -0.9258684
[4,] -0.9370340  0.9763975 -0.9258684  0.9800653
>
```

Geometry of eigenvalues and eigenvectors



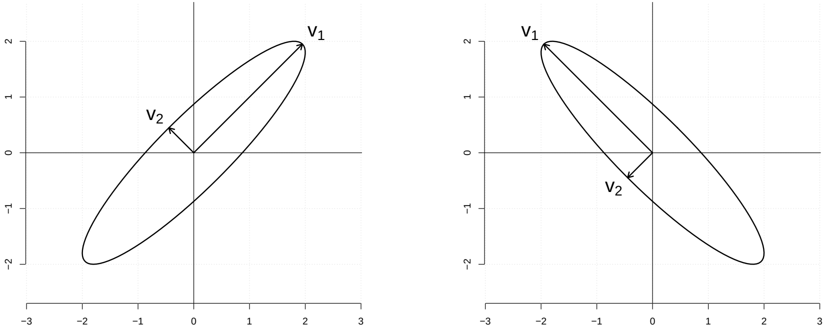
- $x_1^2 + x_2^2 = c^2$ is a circle.
- $a_1 x_1^2 + a_2 x_2^2 = c^2$ and $a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_2 x_2^2 = c^2$ are ellipses.
- Constant (squared) distance from origin: $\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$
- $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'(\lambda_1 \mathbf{v}_1 \mathbf{v}_1' + \lambda_2 \mathbf{v}_2 \mathbf{v}_2')\mathbf{x} = \lambda_1 (\mathbf{x}'\mathbf{v}_1)^2 + \lambda_2 (\mathbf{x}'\mathbf{v}_2)^2 = \lambda_1 \tilde{x}_1^2 + \lambda_2 \tilde{x}_2^2 = c^2$
- Half the length of ith principal axis: $\frac{c}{\sqrt{\lambda_i}}$
- Eigenvectors are principal axes, eigenvalues relate to the length of the principal axes.

Geometry of eigenvalues and eigenvectors

$$\mathbf{R}_1 = \begin{bmatrix} 1.00 & +0.90 \\ +0.90 & 1.00 \end{bmatrix} \qquad \mathbf{R}_2 = \begin{bmatrix} 1.00 & -0.90 \\ -0.90 & 1.00 \end{bmatrix}$$

$$\lambda_1 = 1.9, \lambda_2 = 0.1 \qquad \lambda_1 = 1.9, \lambda_2 = 0.1$$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = 1 \text{ with } \mathbf{A} = \mathbf{R}_1^{-1} \qquad \mathbf{x}'\mathbf{A}\mathbf{x} = 1 \text{ with } \mathbf{A} = \mathbf{R}_2^{-1}$$



Generalization

- The spectral decomposition allows to approximate a square symmetric matrix.
- Mostly used in those multivariate methods that approximate covariance or correlation matrices.
- It is (often) of interest to approximate just any rectangular matrix.
- This can be done by the singular value decomposition (SVD).
- Also known as the Eckart-Young theorem (1936).
- The spectral decomposition is a special case of the SVD.

Singular value decomposition (compact)

Any real matrix $n \times p$ matrix $\mathbf{X}$ can be decomposed as

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

- $\mathbf{U}$ $n \times r$ matrix of orthonormal left singular vectors. $\mathbf{U}'\mathbf{U} = \mathbf{I}_r$
- $\mathbf{D}$ $r \times r$ diagonal matrix of non-increasing positive singular values $(d_{11} \geq d_{22} \geq \cdots \geq d_{rr})$.
- $\mathbf{V}$ $p \times r$ matrix of orthonormal right singular vectors. $\mathbf{V}'\mathbf{V} = \mathbf{I}_r$

Alternatively

$$\mathbf{X} = \sum_{i=1}^{r} d_{ii}\mathbf{u}_i\mathbf{v}_i' = d_1\mathbf{u}_1\mathbf{v}_1' + d_2\mathbf{u}_2\mathbf{v}_2' + \cdots + d_r\mathbf{u}_r\mathbf{v}_r'$$

Singular value decompostion (theorem)

A rank k approximation $\hat{\mathbf{X}}$ to matrix $\mathbf{X}$, optimal in the least squares sense, is obtained as

$$\hat{\mathbf{X}} = \mathbf{U}_{[,1:k]}\mathbf{D}_{[1:k,1:k]}\mathbf{V}_{[,1:k]}{}'$$

E.g., a rank 2 approximation to matrix $\mathbf{X}$ is obtained by

$$\mathbf{U}_{n\times 2}\mathbf{D}_{(2\times 2)}\mathbf{V}_{p\times 2}{}'$$

Numerical example: allele frequencies of Y-STR DYS448 over 129 populations worldwide

Original data

Population	A19	A20	A21	Other
Arg-B	0.402	0.337	0.109	0.152
Arg-F	0.423	0.310	0.183	0.085
Arg-M	0.446	0.337	0.119	0.099
Arg-N	0.460	0.280	0.060	0.200
Arg-S	0.500	0.200	0.120	0.180
Aus	0.417	0.386	0.093	0.104
Bel-A	0.558	0.296	0.058	0.087
Bel-V	0.600	0.305	0.029	0.067
Ben	0.078	0.137	0.627	0.157
Bol-M	0.432	0.409	0.091	0.068
Bol-N	0.286	0.589	0.089	0.036
Bos	0.440	0.460	0.060	0.040
Bra-R	0.431	0.228	0.195	0.146
Bra-SG	0.574	0.197	0.164	0.066
Bra-SP	0.450	0.275	0.200	0.075
Chi-B	0.370	0.325	0.110	0.195
Chi-C	0.440	0.190	0.060	0.310
Chi-Sh	0.321	0.450	0.055	0.174
Chi-So	0.567	0.133	0.033	0.267
Chi-Xi	0.065	0.576	0.043	0.315
Chi-Xu	0.366	0.255	0.028	0.352
Chi-Y	0.307	0.386	0.030	0.277
CoR	0.452	0.253	0.163	0.133
Cro-C	0.432	0.456	0.064	0.048
Cro-Z	0.351	0.553	0.088	0.009
Cze-B	0.347	0.625	0.014	0.014
Cze-M	0.310	0.524	0.119	0.048
Den	0.411	0.497	0.054	0.038
ENG-C	0.568	0.296	0.037	0.099
ENG-S	0.579	0.263	0.053	0.105
⋮	⋮	⋮	⋮	⋮
Wal	0.720	0.144	0.025	0.110
Zim	0.055	0.109	0.764	0.073

Rank 2 approximation

Population	A19	A20	A21	Other
Arg-B	0.405	0.335	0.111	0.149
Arg-F	0.376	0.344	0.139	0.141
Arg-M	0.418	0.357	0.093	0.132
Arg-N	0.464	0.277	0.064	0.195
Arg-S	0.458	0.230	0.080	0.231
Aus	0.414	0.388	0.090	0.108
Bel-A	0.506	0.334	0.008	0.152
Bel-V	0.539	0.349	-0.029	0.141
Ben	0.002	0.193	0.555	0.250
Bol-M	0.419	0.419	0.078	0.084
Bol-N	0.333	0.555	0.135	-0.022
Bos	0.433	0.465	0.053	0.049
Bra-R	0.383	0.262	0.150	0.205
Bra-SG	0.467	0.275	0.062	0.197
Bra-SP	0.383	0.324	0.136	0.157
Chi-B	0.392	0.309	0.131	0.168
Chi-C	0.467	0.171	0.085	0.277
Chi-Sh	0.385	0.403	0.116	0.096
Chi-So	0.544	0.150	0.011	0.295
Chi-Xi	0.264	0.431	0.233	0.072
Chi-Xu	0.444	0.198	0.102	0.256
Chi-Y	0.401	0.318	0.119	0.162
CoR	0.407	0.286	0.120	0.187
Cro-C	0.428	0.459	0.060	0.053
Cro-Z	0.367	0.541	0.103	-0.011
Cze-B	0.397	0.589	0.061	-0.047
Cze-M	0.335	0.505	0.144	0.016
Den	0.419	0.492	0.062	0.028
ENG-C	0.522	0.330	-0.007	0.155
ENG-S	0.522	0.305	-0.002	0.175
⋮	⋮	⋮	⋮	⋮
Wal	0.614	0.222	-0.076	0.241
Zim	-0.080	0.207	0.636	0.237

Purps, J. et al. (2014) A global analysis of Y–chromosomal haplotype diversity for 23 STR loci. Forensic Science International: Genetics 12: 12–23.

Singular vectors are eigenvectors

- $\mathbf{X}'\mathbf{X} = \mathbf{V}\mathbf{D}\mathbf{U}'\mathbf{U}\mathbf{D}\mathbf{V}' = \mathbf{V}\mathbf{D}^2\mathbf{V}'$
- $\mathbf{X}\mathbf{X}' = \mathbf{U}\mathbf{D}\mathbf{V}'\mathbf{V}\mathbf{D}\mathbf{U}' = \mathbf{U}\mathbf{D}^2\mathbf{U}'$
- Eigenvalues of $\mathbf{X}\mathbf{X}'$ and $\mathbf{X}'\mathbf{X}$ are squared singular values.
- Singular vectors are eigenvectors, $\mathbf{U}$ of $\mathbf{X}\mathbf{X}'$ and $\mathbf{V}$ of $\mathbf{X}'\mathbf{X}$.

Singular value decomposition (extended)

Sometimes the svd is also written as

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

- $\mathbf{U}$ $n \times p$ matrix of orthonormal left singular vectors. $\mathbf{U}'\mathbf{U} = \mathbf{I}_p$
- $\mathbf{D}$ $p \times p$ diagonal matrix of non-increasing singular values.
- $\mathbf{V}$ $p \times p$ matrix of orthonormal right singular vectors. $\mathbf{V}'\mathbf{V} = \mathbf{I}_p$

where matrix $\mathbf{D}$ now has trailing zeros on the diagonal.

Singular value decomposition in R

```
> X <- read.table("http://www-eio.upc.es/~jan/SISG/PurpsAlleleFreq.dat",header=TRUE)
> rownames(X) <- X[,1]
> X <- X[,-1]
> head(X)
            A19       A20        A21      Other
Arg-B 0.4021739 0.3369565 0.10869565 0.15217391
Arg-F 0.4225352 0.3098592 0.18309859 0.08450704
Arg-M 0.4455445 0.3366337 0.11881188 0.09900990
Arg-N 0.4600000 0.2800000 0.06000000 0.20000000
Arg-S 0.5000000 0.2000000 0.12000000 0.18000000
Aus   0.4169884 0.3861004 0.09266409 0.10424710
>
> m <- colMeans(X)
> m
      A19       A20       A21     Other
0.3868279 0.3469738 0.1269583 0.1392401
>
> Xc <- scale(X,scale=FALSE)
>
> out <- svd(Xc)
> U <- out$u
> D <- diag(out$d)
> V <- out$v
>
> head(U)
             [,1]         [,2]         [,3]        [,4]
[1,]  0.004474539  0.014019460 -0.003715464  0.15739050
[2,]  0.003501264 -0.007690527  0.064411346  0.15890133
[3,] -0.010391789  0.022202399  0.037869618  0.15633749
[4,]  0.030303555  0.061403042 -0.005661096 -0.01203709
[5,]  0.059008082  0.060551765  0.057607799 -0.00997095
[6,] -0.028691684  0.016890721  0.004138589 -0.01297603
> t(U)%*%U
              [,1]          [,2]          [,3]          [,4]
[1,]  1.000000e+00 -8.326673e-17 -1.040834e-17 -1.595946e-16
[2,] -8.326673e-17  1.000000e+00 -5.551115e-17 -4.857226e-17
[3,] -1.040834e-17 -5.551115e-17  1.000000e+00  3.469447e-17
[4,] -1.595946e-16 -4.857226e-17  3.469447e-17  1.000000e+00
> head(V)
            [,1]       [,2]       [,3] [,4]
[1,] -0.07470358  0.7002366  0.5040714 -0.5
[2,] -0.76460411 -0.1739959 -0.3675677 -0.5
[3,]  0.24997587 -0.6759941  0.4801500 -0.5
[4,]  0.58933182  0.1497533 -0.6166538 -0.5
```

Singular value decomposition in R

```
> t(V)%*%V
              [,1]         [,2]          [,3]          [,4]
[1,]  1.000000e+00 5.551115e-17  2.220446e-16 -5.551115e-17
[2,]  5.551115e-17 1.000000e+00  1.387779e-17  1.110223e-16
[3,]  2.220446e-16 1.387779e-17  1.000000e+00 -5.551115e-17
[4,] -5.551115e-17 1.110223e-16 -5.551115e-17  1.000000e+00
> D
        [,1]     [,2]     [,3]         [,4]
[1,] 2.13876 0.000000 0.000000 0.000000e+00
[2,] 0.00000 1.909569 0.000000 0.000000e+00
[3,] 0.00000 0.000000 1.433727 0.000000e+00
[4,] 0.00000 0.000000 0.000000 2.970001e-08
>
> U2 <- U[,1:2]
> D2 <- D[1:2,1:2]
> V2 <- V[,1:2]
>
> Xhat <- U2%*%D2%*%t(V2)
> head(Xhat)
            [,1]         [,2]        [,3]         [,4]
[1,]  0.01803122 -0.011975303 -0.01570487  0.009648952
[2,] -0.01084280 -0.003170401  0.01179929  0.002213915
[3,]  0.03134828  0.009616835 -0.03421599 -0.006749126
[4,]  0.07726341 -0.069957158 -0.06306114  0.055754880
[5,]  0.07153892 -0.116614967 -0.04661572  0.091691764
[6,]  0.02716959  0.041307586 -0.03714319 -0.031333986
>
> Xhat <- Xhat + rep(1,nrow(X))%o%m
>
> head(X)
            A19       A20        A21      Other
Arg-B 0.4021739 0.3369565 0.10869565 0.15217391
Arg-F 0.4225352 0.3098592 0.18309859 0.08450704
Arg-M 0.4455445 0.3366337 0.11881188 0.09900990
Arg-N 0.4600000 0.2800000 0.06000000 0.20000000
Arg-S 0.5000000 0.2000000 0.12000000 0.18000000
Aus   0.4169884 0.3861004 0.09266409 0.10424710
> head(Xhat)
           A19       A20        A21     Other
[1,] 0.4048591 0.3349985 0.11125339 0.1488890
[2,] 0.3759851 0.3438034 0.13875755 0.1414540
[3,] 0.4181761 0.3565906 0.09274227 0.1324909
[4,] 0.4640913 0.2770166 0.06389712 0.1949949
[5,] 0.4583668 0.2303588 0.08034254 0.2309318
[6,] 0.4139975 0.3882814 0.08981507 0.1079061
```

Goodness-of-fit

- How good (or bad) is our approximation to **X**?
- Some statistic expressing goodness of fit is needed (like R^2 in regression)
- The singular values are informative about the goodness-of-fit

Note that

$$\mathrm{tr}(\mathbf{X}'\mathbf{X}) = \mathrm{tr}(\mathbf{VDU}'\mathbf{UDV}') = \mathrm{tr}(\mathbf{VD}^2\mathbf{V}') = \mathrm{tr}(\mathbf{V}'\mathbf{VD}^2) = \mathrm{tr}(\mathbf{D}^2) = \sum_{j=1}^{p} d_{jj}^2 = \sum_{j=1}^{p} \lambda_j$$

And that for a rank 2 approximation

$$\begin{aligned}\mathrm{tr}(\hat{\mathbf{X}}'\hat{\mathbf{X}}) &= \mathrm{tr}(\mathbf{V}_{[,1:2]}\mathbf{D}_{[1:2,1:2]}\mathbf{U}'_{[,1:2]}\mathbf{U}_{[,1:2]}\mathbf{D}_{[1:2,1:2]}\mathbf{V}'_{[,1:2]}) = \mathrm{tr}(\mathbf{V}_{[,1:2]}\mathbf{D}^2_{[1:2,1:2]}\mathbf{V}'_{[,1:2]}) = \mathrm{tr}(\mathbf{V}'_{[,1:2]}\mathbf{V}_{[,1:2]}\mathbf{D}^2_{[1:2,1:2]})\\ &= \mathrm{tr}(\mathbf{D}^2_{[1:2,1:2]}) = d_{11}^2 + d_{22}^2 = \lambda_1 + \lambda_2\end{aligned}$$

And that for the error matrix

$$\mathrm{tr}(\mathbf{E}'\mathbf{E}) = \mathrm{tr}((\mathbf{X} - \hat{\mathbf{X}})'(\mathbf{X} - \hat{\mathbf{X}})) = \mathrm{tr}(\mathbf{V}_{[,3:p]}\mathbf{D}^2_{[3:p,3:p]}\mathbf{V}'_{[,3:p]}) = \lambda_3 + \lambda_4 + \cdots \lambda_p$$

And a natural measure for goodness-of-fit is

$$\frac{\mathrm{tr}(\hat{\mathbf{X}}'\hat{\mathbf{X}})}{\mathrm{tr}(\mathbf{X}'\mathbf{X})} = \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^{p} \lambda_j}$$

Similar to the total, explained and residual sum-of-squares in regression.

Weighted singular value decomposition

- On occasions we may wish to use weights for cases (rows, r_i) and/or variables (columns, c_j)
- We normally minimize $\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 = \mathrm{tr}(\mathbf{E}'\mathbf{E})$
- Define $\mathbf{D}_r$ with weights for the rows $\mathbf{D}_c$ with weights for the columns.
- We now wish to minimize $\sum_{i=1}^{n} \sum_{j=1}^{p} r_i c_j e_{ij}^2 = \mathrm{tr}(\mathbf{D}_c \mathbf{E}' \mathbf{D}_r \mathbf{E})$
- Note that $\sum_{i=1}^{n} \sum_{j=1}^{p} r_i c_j e_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\sqrt{r_i}\sqrt{c_j} e_{ij}\right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{e}_{ij}^2$
- $\sum_{i=1}^{n} \sum_{j=1}^{p} \tilde{e}_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\sqrt{r_i}\sqrt{c_j} x_{ij} - \sqrt{r_i}\sqrt{c_j} \hat{x}_{ij}\right)^2$
- Solution obtained by transforming the data prior to the svd, and backtransforming afterwards

$$\mathbf{X}_t = \mathbf{D}_r^{\frac{1}{2}} \mathbf{X} \mathbf{D}_c^{\frac{1}{2}} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

  Now compute $\tilde{\mathbf{U}} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U}$ and $\tilde{\mathbf{V}} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V}$
- Note that $\tilde{\mathbf{U}}\mathbf{D}\tilde{\mathbf{V}}' = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U}\mathbf{D}\mathbf{V}' \mathbf{D}_c^{-\frac{1}{2}} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{D}_r^{\frac{1}{2}} \mathbf{X} \mathbf{D}_c^{\frac{1}{2}} \mathbf{D}_c^{-\frac{1}{2}} = \mathbf{X}$
- $\tilde{\mathbf{U}}_{[,1:k]} \mathbf{D}_{[1:k,1:k]} \tilde{\mathbf{V}}'_{[,1:k]}$ is a rank k approximation to $\mathbf{X}$ in the weighted least squares sense.

References

- Johnson & Wichern, (2002) *Applied Multivariate Statistical Analysis*, 5th edition, Prentice Hall, Chapter 2.
- Mardia, K.V. et al. (1979) *Multivariate Analysis*. Academic press. Appendix A.