Singular Value Decomposition

Jan Graffelman[1]

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

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

jan.graffelman@upc.edu

February 13, 2018

A data matrix

	Husband age	Husband height	Wife age	Wife height
1	49	1809	43	1590
2	25	1841	28	1560
3	40	1659	30	1620
4	52	1779	57	1540
5	58	1616	52	1420
6	32	1695	27	1660
7	43	1730	52	1610
8	47	1740	43	1580
9	31	1685	23	1610
10	26	1735	25	1590

Age and height of husband and wife for 10 couples

The problem

- The data matrix $\mathbf{X}$ is 10×4, and of rank 4.
- Can we approximate $\mathbf{X}$ by a rank 2 matrix, say $\hat{\mathbf{X}}$
- Entries of $\hat{\mathbf{X}}$ must be as "close" as possible to $\mathbf{X}$
- Note: a rank 2 matrix can be represented in a two-dimensional graph.

The Solution

$$\mathbf{X} = \begin{bmatrix} 49 & 1809 & 43 & 1590 \\ 25 & 1841 & 28 & 1560 \\ 40 & 1659 & 30 & 1620 \\ 52 & 1779 & 57 & 1540 \\ 58 & 1616 & 52 & 1420 \\ 32 & 1695 & 27 & 1660 \\ 43 & 1730 & 52 & 1610 \\ 47 & 1740 & 43 & 1580 \\ 31 & 1685 & 23 & 1610 \\ 26 & 1735 & 25 & 1590 \end{bmatrix} \quad \hat{\mathbf{X}} = \begin{bmatrix} 43.94 & 1809.21 & 43.72 & 1589.89 \\ 46.36 & 1838.22 & 48.00 & 1562.03 \\ 35.09 & 1659.32 & 29.27 & 1619.79 \\ 44.07 & 1780.44 & 44.75 & 1538.92 \\ 39.40 & 1618.03 & 39.32 & 1418.55 \\ 35.64 & 1694.60 & 29.49 & 1660.28 \\ 39.24 & 1731.53 & 36.00 & 1608.81 \\ 40.67 & 1740.69 & 38.73 & 1579.51 \\ 36.67 & 1683.96 & 31.92 & 1610.78 \\ 39.90 & 1733.25 & 37.32 & 1591.27 \end{bmatrix}$$

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 $\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}$.
- The least squares criterion amounts to $\sum_{i=1}^{n} \sum_{j=1}^{p} e_{ij}^2 = \mathrm{tr}(\mathbf{E}'\mathbf{E})$.

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

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("D:/data/HusbandsAndWives.dat")
X <- read.table("http://www-eio.upc.edu/~jan/data/MVA/HusbandsAndWives.dat")

X <- as.matrix(X)
X <- X[,1:4]

svd.results <- svd(X)
U <- svd.results$u
V <- svd.results$v
D <- diag(svd.results$d)
print(U)
print(V)
print(D)

U2 <- U[,1:2]
V2 <- V[,1:2]
D2 <- D[1:2,1:2]

Xhat <- U2%*%D2%*%t(V2)
print(Xhat)
```

Goodness-of-fit

- How good (or bad) is our approximation to $\mathbf{X}$?
- 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

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

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 and/or variables.
- $\mathbf{D}_{w_r}$ weights for the rows $\mathbf{D}_{w_c}$ weights for the columns.
- We now wish to minimize $\sum_{i=1}^{n} \sum_{j=1}^{p} w_{r_i} w_{c_j} e_{ij}^2 = \mathrm{tr}(\mathbf{D}_{w_c} \mathbf{E}' \mathbf{D}_{w_r} \mathbf{E})$
- Solution obtained by transforming the data prior to the svd

$$\mathbf{X}_t = \mathbf{D}_{w_r}^{\frac{1}{2}} \mathbf{X} \mathbf{D}_{w_c}^{\frac{1}{2}} = \mathbf{U}\mathbf{D}\mathbf{V}'$$

  Now compute $\tilde{\mathbf{U}} = \mathbf{D}_{w_r}^{-\frac{1}{2}} \mathbf{U}$ and $\tilde{\mathbf{V}} = \mathbf{D}_{w_c}^{-\frac{1}{2}} \mathbf{V}$

- Note that $\tilde{\mathbf{U}}\mathbf{D}\tilde{\mathbf{V}}' = \mathbf{D}_{w_r}^{-\frac{1}{2}} \mathbf{U}\mathbf{D}\mathbf{V}' \mathbf{D}_{w_c}^{-\frac{1}{2}} = \mathbf{D}_{w_r}^{-\frac{1}{2}} \mathbf{D}_{w_r}^{\frac{1}{2}} \mathbf{X} \mathbf{D}_{w_c}^{\frac{1}{2}} \mathbf{D}_{w_c}^{-\frac{1}{2}} = \mathbf{X}$
- $\tilde{\mathbf{U}}_{[,1:k]} \mathbf{D}_{[1:k,1:k]} \tilde{\mathbf{V}}'_{[,1:k]}$ 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.
- Greenacre, M. (2010) Biplots in Practice. Chapter 5.