CS724: Topics in Algorithms
Biplots

Prof. Dan A. Simovici

Biplots introduced by K. R. Gabriel offer a succinct and powerful way of representing graphically the elements of a data matrix using two sets of vectors that represent the objects and the variables (hence, the term *biplot*). We shall assume that there are more objects than variables, so for the data matrix $A \in \mathbb{R}^{m \times n}$ we have $m > n$.

Suppose that $A \in \mathbb{R}^{m \times n}$ can be written as a product, $A = LR$, where $L \in \mathbb{R}^{m \times r}$, $R \in \mathbb{R}^{r \times n}$ are the *left* and the *right* factors, respectively,

$$L = \begin{pmatrix} \mathbf{l}'_1 \\ \vdots \\ \mathbf{l}'_m \end{pmatrix} \text{ and } R = (\mathbf{r}_1 \ \cdots \ \mathbf{r}_n)$$

where $\{\mathbf{l}_1, \ldots, \mathbf{l}_m\} \subset \mathbb{R}^r$ and $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\} \subset \mathbb{R}^r$.

Each element a_{ij} of A can be regarded as a scalar product of two vectors

$$a_{ij} = \mathbf{l}'_i \mathbf{r}_j$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Examples of LR factorization of matrices

- the *Cholesky factorization* of symmetric positive definite matrices as a product $A = R'R$, where R is a unique upper triangular matrix R with positive diagonal elements;
- the *thin QR factorization* for a full-rank matrix $A \in \mathbb{R}^{m \times n}$ as $A = QR$ where $Q \in \mathbb{R}^{m \times n}$ and $R \in \mathbb{R}^{n \times n}$, where the columns of the orthogonal matrix Q form an orthonormal basis for Span(A), and $R = (r_{ij})$ is an upper triangular invertible matrix such that its diagonal elements are non-negative numbers;
- the *full QR factorization* for a full-rank matrix $A \in \mathbb{R}^{m \times n}$ with $m > n$ as

$$A = Q \begin{pmatrix} R \\ O_{m-n,n} \end{pmatrix},$$

where $Q \in \mathbb{R}^{m \times m}$ is an orthogonal matrix and $R \in \mathbb{R}^{n \times n}$ is an upper triangular matrix having non-negative diagonal entries.

Cholesky Factorization in **R**

```
> A
     [,1] [,2] [,3]
[1,]    2   -1    0
[2,]   -1    2   -1
[3,]    0   -1    2
> chol(A)
         [,1]       [,2]       [,3]
[1,] 1.414214 -0.7071068  0.0000000
[2,] 0.000000  1.2247449 -0.8164966
[3,] 0.000000  0.0000000  1.1547005
```

```
> d <- qr(A)
```

The factors can be determined as

```
> Q <- qr.Q(d)
> R <- qr.R(d)
> Q
           [,1]       [,2]      [,3]
[1,] -0.8944272 -0.3585686 0.2672612
[2,]  0.4472136 -0.7171372 0.5345225
[3,]  0.0000000  0.5976143 0.8017837
> R
          [,1]      [,2]       [,3]
[1,] -2.236068  1.788854 -0.4472136
[2,]  0.000000 -1.673320  1.9123658
[3,]  0.000000  0.000000  1.0690450
```

The result can be retrieved as

```
> Q %*% R
     [,1] [,2]          [,3]
[1,]    2   -1 -4.440892e-16
[2,]   -1    2 -1.000000e+00
[3,]    0   -1  2.000000e+00
```

- LR factorization is not unique. Starting from the factorization $A = LR$ new factorizations of A can be built as $A = (LK')(R'K^{-1})'$ for every invertible matrix $K \in \mathbb{R}^{r \times r}$.
- To use the biplot for a representation of the relations between the rows $\mathbf{w}_1, \ldots, \mathbf{w}_m$ of A one could choose R such that $RR' = I_r$, which yields $AA' = LL'$. This implies $\mathbf{w}_i'\mathbf{w}_j = \mathbf{l}_i'\mathbf{l}_j$ for $1 \leq i, j \leq m$. Taking $i = j$ we have $\| \mathbf{w}_i' \| = \| \mathbf{l}_i' \|$, which , in turn, implies $\angle(\mathbf{w}_i', \mathbf{w}_j') = \angle(\mathbf{l}_i', \mathbf{l}_j')$.
- A similar choice can be made for the columns of A by imposing the requirement $L'L = I_r$, which implies $A'A = R'R$.

The case when the rank r of the matrix A is 2 is especially interesting because we can draw the vectors $\mathbf{l}_1, \ldots, \mathbf{l}_m, \mathbf{r}_1, \ldots, \mathbf{r}_n$ to obtain an exact two-dimensional representation of A, as we show in the next example.

Example

Let

$$A = \begin{pmatrix} 18 & 8 & 20 \\ -4 & 20 & 1 \\ 25 & 8 & 27 \\ 9 & 4 & 10 \end{pmatrix}$$

be a matrix of rank 2 in $\mathbb{R}^{4\times 3}$ that can be written as $A = LR$, where

$$L = \begin{pmatrix} 2 & 4 \\ -2 & 3 \\ 3 & 5 \\ 1 & 2 \end{pmatrix} \text{ and } R = \begin{pmatrix} 5 & -4 & 4 \\ 2 & 4 & 3 \end{pmatrix}.$$

The vectors that help us with the representation of A are

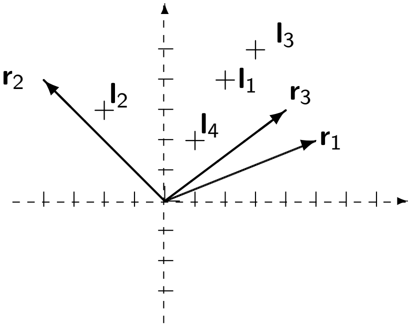
$$\mathbf{l}_1 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \mathbf{l}_2 = \begin{pmatrix} -2 \\ 3 \end{pmatrix}, \mathbf{l}_3 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \mathbf{l}_4 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and

$$\mathbf{r}_1 = \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \mathbf{r}_2 = \begin{pmatrix} -4 \\ 4 \end{pmatrix}, \mathbf{r}_3 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

For example, the a_{32} element of A can be written as

$$a_{32} = \mathbf{l}_3' \mathbf{r}_2 = (3\ 5) \begin{pmatrix} -4 \\ 4 \end{pmatrix} = 8.$$



Representation of the vectors $\mathbf{l}_i$ and $\mathbf{r}_j$

Each vector $\mathbf{l}_i$ corresponds to a row of A and each vector $\mathbf{r}_j$ to a column of A.

When we can factor a sample data matrix X as $X = LR$ a column $\mathbf{r}_j$ of the right factor is referred to as the *biplot axis* and corresponds to a variable $\mathcal{V}_j$.

Each vector $\mathbf{l}_i'$ represents an observation in the sample matrix. It is interesting to note that the magnitude of projection of $\mathbf{l}_i$ on the biplot axis $\mathbf{r}_j$ is

$$\parallel \mathbf{l}_i \parallel_2 \cos\angle(\mathbf{l}_i, \mathbf{r}_j) = \frac{\mathbf{l}_i'\mathbf{r}_j}{\parallel \mathbf{r}_j \parallel_2} = \frac{a_{ij}}{\parallel \mathbf{r}_j \parallel_2}.$$

Therefore, if we choose $\frac{1}{\|\mathbf{r}_j\|_2}$ as the unit of measure on the axis $\mathbf{r}_j$ we can read the values of the entries a_{ij} directly on the axis $\mathbf{r}_j$. For instance, the unit along the biplot axis is $\frac{1}{\|\mathbf{r}_3\|_2} = 0.2$. It is also clear that if two axis of the biplot point roughly in the same direction, the corresponding variables will show a strong correlation.

In general, the rank of the data matrix A is larger than 2. In this case, approximative representations of A can be obtained by using the thin singular value decomposition of matrices.
Let A be a matrix of rank r and let

$$A = UDV' = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i',$$

be the thin SVD, where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ are matrices of rank r (and, therefore, full-rank matrices) having orthonormal sets of columns.
Here $U = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_r)$ and $V = (\mathbf{v}_1 \ \cdots \ \mathbf{v}_r)$.

The matrix D containing singular values can be split between U and V by defining $L = U\sqrt{D}$ and $R = \sqrt{D}V'$. By Eckhart-Young Theorem the best approximation of A in the sense of the matrix norm $\| \cdot \|_2$ in the class of matrix of rank k is the matrix defined by

$$B(k) = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i'.$$

The same matrix $B(k)$ is the best approximation of A in the sense of Frobenius norm. The extent of the defficiency of this approximation is measured by $\| A - B(k) \|_F^2 = \sigma_{k+1}^2 + \cdots + \sigma_r^2$. Since $\| A \|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2$, an absolute measure of the quality of the approximation of A by $B(k)$ is

$$q_k = 1 - \frac{\| A - B(k) \|_F^2}{\| A \|_F^2} = \frac{\sigma_1^2 + \cdots + \sigma_k^2}{\sigma_1^2 + \cdots + \sigma_r^2}$$

In the special case, $k = 2$, the quality of the approximation is

$$q_2 = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \cdots + \sigma_r^2}$$

and it is desirable that this number is as close as one as possible. The rank-2 approximation of A is useful because we can apply biplots to the visualization of A.

Let $A \in \mathbb{R}^{5\times 3}$ be the matrix defined by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to see that the rank of this matrix is 3.

A singular value decomposition can be obtained using (svd(A)) which yields

```
$d
[1] 2.358294 1.199353 1.000000

$u
           [,1]       [,2]          [,3]
[1,] -0.2786727  0.2175811  7.071068e-01
[2,] -0.2786727  0.2175811 -7.071068e-01
[3,] -0.7138349 -0.3397643 -9.677906e-17
[4,] -0.5573454  0.4351621  3.420751e-16
[5,] -0.1564894 -0.7749265 -4.388542e-16

$v
           [,1]       [,2]          [,3]
[1,] -0.6571923  0.2609565  7.071068e-01
[2,] -0.6571923  0.2609565 -7.071068e-01
[3,] -0.3690482 -0.9294103 -4.996004e-16
```

The rank-2 approximation of this matrix is:

$$B(2) = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathsf{H}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathsf{H}},$$

and is computed in **R** using

```
> B2 <- 2.3583 * u[,1] %*% t(v[,1]) + 1.1994 * u[,2] %*% t(v[,2])

B2 =
     0.5000     0.5000    -0.0000
     0.5000     0.5000    -0.0000
     1.0000     1.0000     1.0000
     1.0000     1.0000    -0.0000
    -0.0000    -0.0000     1.0000
```

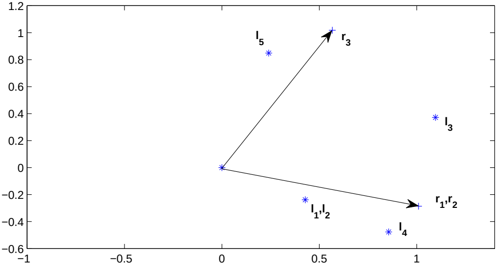
If we split the singular values as

$$B(2) = (\sqrt{\sigma_1}\mathbf{u}_1)(\sqrt{\sigma_1}\mathbf{v}_1)^{\mathsf{H}} + (\sqrt{\sigma_2}\mathbf{u}_2)(\sqrt{\sigma_2}\mathbf{v}_2)^{\mathsf{H}},$$

then $B(2)$ can be written as

$$B(2) = \begin{pmatrix} 0.4280 & -0.2383 \\ 0.4280 & -0.2383 \\ 1.0962 & 0.3721 \\ 0.8559 & -0.4766 \end{pmatrix} \begin{pmatrix} 1.0092 & 1.0092 & 0.5667 \\ -0.2858 & -0.2858 & 1.0179 \end{pmatrix}.$$

The biplot of rank 2 approximation of the matrix A:



The quality of the approximation of A is

$$q_2 = \frac{2.3583^2 + 1.1994^2}{2.3583^2 + 1.1994^2 + 1} = 0.875$$

The "allocation" of singular values among the columns of the matrices U and V may lead to biplots that have distinct properties.
For example, we could write

$$B(2) = (\sigma_1 \mathbf{u}_1)\mathbf{v}_1^{\mathsf{H}} + (\sigma_2 \mathbf{u}_2)\mathbf{v}_2^{\mathsf{H}},$$

or

$$B(2) = \mathbf{u}_1(\sigma_1 \mathbf{v}_1)^{\mathsf{H}} + \mathbf{u}_2(\sigma_2 \mathbf{v}_2)^{\mathsf{H}}.$$

The first allocation leads to the factorization $B(2) = LR$, where

$$L = \begin{pmatrix} 0.6572 & -0.2610 \\ 0.6572 & -0.2610 \\ 1.6834 & 0.4075 \\ 1.3144 & -0.5219 \\ 0.3690 & 0.9294 \end{pmatrix} \text{ and } R = \big(0.65720.65720.3690 - 0.2610 - 0.26100$$

while the second yields the factors

$$L = \big(0.2787 \quad -0.21760.2787 \quad -0.21760.7138 \quad 0.33980.5573 \quad -0.43520.15$$

The first variant leads to a representation, where the distances between the vectors $\mathbf{l}_i$ approximate the Euclidean distances between rows, while for the second variant, the cosine of angles between the vectors $\mathbf{r}_j$ approximate the correlations between variables.