Scientific Computing for Biologists

Singular Value Decomposition and Biplots

Paul M. Magwene

Overview of Lecture

- Singular Value Decomposition
  - Algebra of SVD
  - Geometry of SVD
  - Relationship to Eigendecomposition
  - Applications of SVD
- Biplots
  - Simultaneous representation of rows and columns of a matrix

Hands-on Session

- SVD and Biplots in R
- Applications of SVD in R
  - ‘Seriation’ using SVD
  - Matrix approximation and image compression using SVD

Matrix Decomposition

- A “factorization” is a re-writing of a mathematical object (number, function, etc.) into the product of other object.
- Your familiar with factorization as used in arithmetic and algebra.

$$12 = 3 \times 4 = 2 \times 6$$
$$x^2 - 4 = (x - 2)(x + 2)$$

Matrix decomposition is the same idea! A matrix decomposition is a factoring of a matrix into simpler parts.

Eigendecomposition

You've already been introduced to one way to decompose a square matrix, $\boldsymbol{A}$:

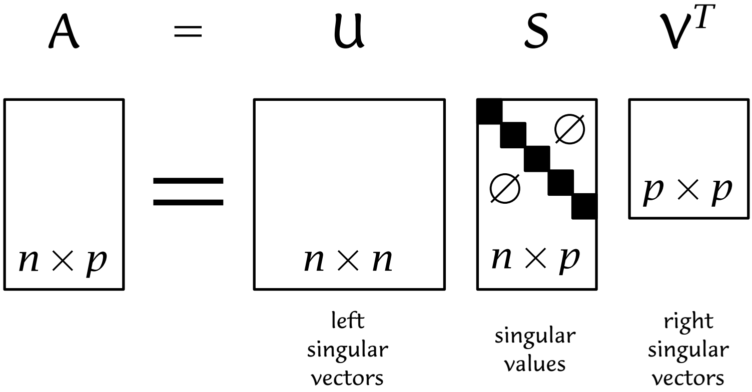
$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$$

where:

- $\boldsymbol{V}$ is a matrix of eigenvectors (in columns)
- $\boldsymbol{D}$ is a diagonal matrix with eigenvalues along diagonal.

when $\boldsymbol{A}$ is real-valued and symmetric than $\boldsymbol{V}$ is orthgonal.

Singular Value Decomposition



- When written like this U and V are orthonormal.
- Sometimes SVD is written as:

$$\underset{(n\times p)}{\boldsymbol{A}} = \underset{(n\times p)}{\boldsymbol{U}}\,\underset{(p\times p)}{\boldsymbol{S}}\,\underset{(p\times p)}{\boldsymbol{V}^T}$$

Facts about SVD

- Singular Value Decomposition is often referred to as giving the "basic structure" of a matrix
- The rank of $\boldsymbol{A}$ is equivalent to the number of non-zero singular values in $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$

$$\text{rank}(\boldsymbol{A}) \leq \min(n, p)$$

- The Euclidean norm (L_2) norm of a matrix is the relative amount it stretches a vector:

$$|\boldsymbol{A}|_E = \frac{|\boldsymbol{A}\boldsymbol{x}|}{|\boldsymbol{x}|}$$

The L_2 norm of $\boldsymbol{A}$ is given by $\boldsymbol{S}_{11}$.

Geometric Interpretation of SVD

Any matrix, $\boldsymbol{A}_{n\times p}$, represents a linear transformation from $\mathbb{R}^p \mapsto \mathbb{R}^n$.

SVD can be thought of decomposing the transformation specified by $\boldsymbol{A}$ into a simple form:

$$\boldsymbol{A} = (\text{rotation})(\text{scaling})(\text{rotation})$$

- $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal (orthogonal) matrices $\rightsquigarrow$ Orthonormal matrices represent rigid rotations (or rotation plus reflection).
- Diagonal matrices represent “stretching”

SVD Example

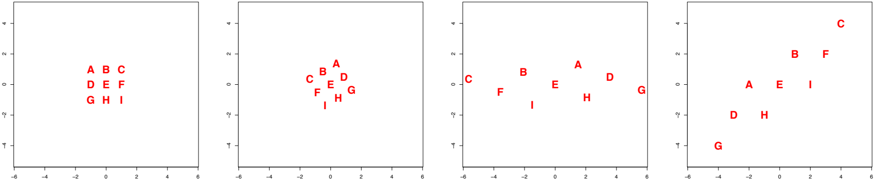
$$A = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix} = USV^T$$

where

$$U = \begin{bmatrix} -0.75 & -0.66 \\ -0.66 & 0.75 \end{bmatrix}, \quad S = \begin{bmatrix} 4.13 & 0 \\ 0 & 0.97 \end{bmatrix}, \quad V^T = \begin{bmatrix} -0.86 & -0.50 \\ 0.50 & -0.86 \end{bmatrix}$$

Geometry

$$x \xrightarrow[\text{(+ reflection)}]{\text{rotation}} V^T x \xrightarrow{\text{scaling}} SV^T x \xrightarrow[\text{(+ reflection)}]{\text{rotation}} USV^T x$$



Relationship of SVD to Eigendecomposition

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T && (1) \\ \boldsymbol{A}^T\boldsymbol{A} &= (\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^T)(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T) && (2) \\ &= \boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T && (3) \\ &= \boldsymbol{V}\boldsymbol{S}\boldsymbol{S}\boldsymbol{V}^T && (4) \end{aligned}$$

Equation 4 follows from the fact that $\boldsymbol{U}$ is orthonormal ($\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$)

If we let $\boldsymbol{D} = \boldsymbol{S}\boldsymbol{S}$, we can rewrite equation 4 as:

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V} \qquad \text{Eigendecomposition!} \qquad (5)$$

- The singular values $\boldsymbol{S}_{ii}$ are $\sqrt{\boldsymbol{D}_{ii}}$ where $\boldsymbol{d}_{ii}$ are the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$.
- The columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$.

Using SVD to do PCA

Let $\boldsymbol{X}$ be a mean-centered $n \times p$ data matrix. The covariance matrix is given by:

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X}$$

By SVD we can write $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, therefore:

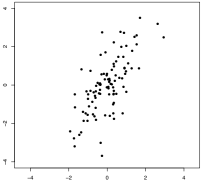
$$\begin{aligned} \boldsymbol{C} &= \frac{1}{n-1}\boldsymbol{V}\boldsymbol{S}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \\ &= \frac{1}{n-1}\boldsymbol{V}\boldsymbol{S}\boldsymbol{S}\boldsymbol{V}^T \end{aligned}$$

- The PC vectors are given by the columns of $\boldsymbol{V}$ (rows of $\boldsymbol{V}^T$)
- The PC scores are given by $\boldsymbol{U}\boldsymbol{D}$, where $\boldsymbol{D} = \boldsymbol{S}\boldsymbol{S}$

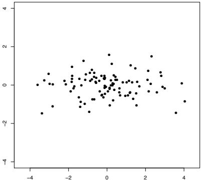
Another Way of Thinking about SVD

Observations in measurement space = Observations in PC Space $\xleftarrow{\text{rotation}}$ Inverse of Eigenvectors

$$X = US\,V^T$$



=



$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

where $\theta \approx \frac{\pi}{3}$

Applications of SVD

- Pseudoinverse of an arbitrary matrix
- Matrix approximation
- Biplots
- ... and many more ...

Pseudoinverse via SVD

The pseudoinverse of a matrix is a generalization of the concept of a matrix inverse. Only square matrices have a matrix inverse; the pseudoinverse applies to an arbitrary $n \times p$ matrix.

Given an $n \times p$ matrix $\boldsymbol{A}$ find matrix $\boldsymbol{A}^+$ such that:

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} &= \boldsymbol{A} \\
\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ &= \boldsymbol{A}^+ \\
(\boldsymbol{A}\boldsymbol{A}^+)^T &= \boldsymbol{A}\boldsymbol{A}^+ \\
(\boldsymbol{A}^+\boldsymbol{A})^T &= \boldsymbol{A}^+\boldsymbol{A}
\end{aligned}$$

Moore-Penrose Inverse via SVD:

$$\begin{aligned}
\text{if } \boldsymbol{A} &= \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \\
\boldsymbol{A}^+ &= \boldsymbol{V}\boldsymbol{S}^+\boldsymbol{U}^T
\end{aligned}$$

where $\boldsymbol{S}^+$ has the reciprocal of non-zero elements of S.

SVD for Matrix Approximation

If $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ then the optimal (least-squares) k-dimensional approximation of $\boldsymbol{A}$ (where $k < \text{rank}(\boldsymbol{A})$) is given by:

$$\tilde{\boldsymbol{A}} = \boldsymbol{U}\boldsymbol{S}^\star\boldsymbol{V}^T$$

where:

$$\begin{aligned} \boldsymbol{S}^\star_{ii} &= \boldsymbol{S}_{ii} \text{ for } i \leq k \\ \boldsymbol{S}^\star_{ii} &= 0 \text{ for } i > k \end{aligned}$$

Biplots

- Technique for simultaneously displaying row and column data (observations and variables)
- Invented by K. Gabriel (see also papers by Gower)

Given a data matrix, $\boldsymbol{X}$, we can use SVD to approximate $\boldsymbol{X}$ as so:

$$\tilde{\boldsymbol{X}}_k = \boldsymbol{U}\boldsymbol{S}^\star\boldsymbol{V}^T$$

(k-dimensional approximation to $\boldsymbol{X}$)

We can rewrite $\tilde{\boldsymbol{X}}_k$ as a product of two matrices:

$$\boldsymbol{X} = \boldsymbol{G}\boldsymbol{H}^T$$

where

$$\boldsymbol{G} = \boldsymbol{U}(\boldsymbol{S}^\star)^\alpha \ \text{ and } \ \boldsymbol{H}^T = (\boldsymbol{S}^\star)^{1-\alpha}\boldsymbol{V}^T$$

Biplots, cont.

$$
\begin{aligned}
\boldsymbol{G} &= \boldsymbol{U}(\boldsymbol{S}^\star)^\alpha \text{ (row effects)} \\
\boldsymbol{H}^T &= (\boldsymbol{S}^\star)^{1-\alpha}\boldsymbol{V}^T \text{ (columns effects)}
\end{aligned}
$$

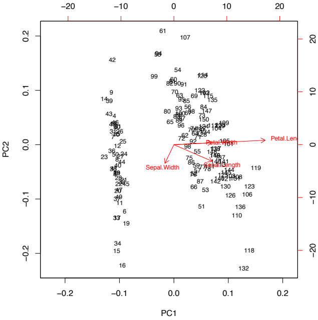
Different choices of α emphasize different relationships in the data.

- $\alpha = 0$, column-metric preserving biplot; optimally approximates variance-covariance structure. Cosine of angles between vectors approximate correlations; distances between points approximate Mahalanobis distance ("correlation biplot")
- $\alpha = 1$, row-metric preserving biplot; optimally approximates Euclidean distances among observations. Coordinates of observations correspond to PC scores; coordinates of variables correspond to eigenvector coefficients ("distance biplot"). PCs are 'sphered'.

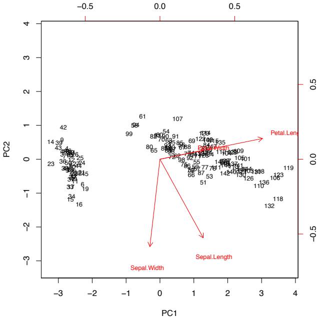
Biplots, Example

- Observations drawn as points in space of PCs
- Variables drawn as vectors in PC space

$\alpha = 0$



$\alpha = 1$



PCA Biplots of Iris data set.