Scientific Computing for Biologists

Singular Value Decomposition and Biplots

Instructor: Paul M. Magwene

02 October 2012

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 matrix decomposition is a factoring of a matrix into simpler parts.

Some familiar factorizations:

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

Matrix decomposition is the same idea!

Eigendecomposition

You've already been introduced to one way to decompose a square matrix, A:

$$A = VDV^{-1}$$

where:

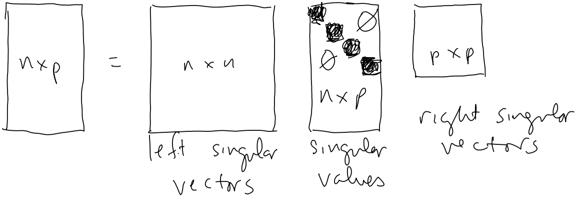
- V is a matrix of eigenvectors (in columns)
- D is a diagonal matrix with eigenvalues along diagonal.

when A is real-valued and symmetric than V is orthgonal.

Singular Value Decomposition

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

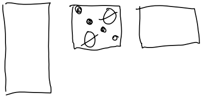
assume $n \geq p$



when written like this U & V are orthonormal

sometimes written as

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



Facts about SVD

- Singular Value Decomposition is often referred to as giving the "basic structure" of a matrix
- The rank of A is equivalent to the number of non-zero singular values in $A = USV^T$

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

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

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

The L_2 norm of A is given by S_{11}.

Geometric Interpretation of SVD

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

SVD can be thought of decomposing the transformation specified by A into a simple form:

$$A = \left(\text{rotation}\right)\left(\text{scaling}\right)\left(\text{rotation}\right)$$

- U and V are orthonormal matrices $\rightsquigarrow$ Orthonormal matrices represent rigid rotations (or rotation plus reflection)
- Diagonal matrices represent "stretching"

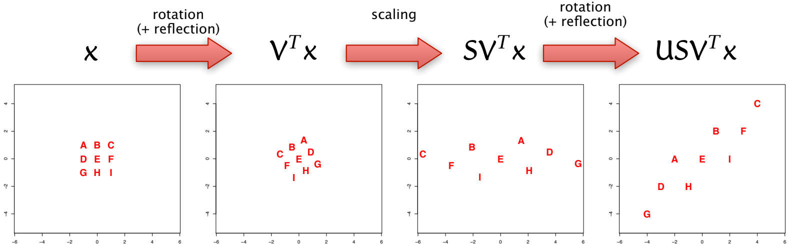
SVD Example

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

where

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

Geometry



Relationship of SVD to Eigendecomposition

$$A = USV^T \tag{1}$$
$$A^T A = (VSU^T)(USV^T) \tag{2}$$
$$= VSU^T USV^T \tag{3}$$
$$= VSSV^T \tag{4}$$

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

If we let $D = SS$, we can rewrite equation 4 as:

$$A^T A = VDV \qquad \text{Eigendecomposition!} \tag{5}$$

- The singlular values S_{ii} are $\sqrt{D_{ii}}$ where d_{ii} are the eigenvalues of $A^T A$.
- The columns of V are the eigenvectors of $A^T A$.

Using SVD to do PCA

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

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

By SVD we can write $X = USV^T$, therefore:

$$\begin{aligned} C &= \frac{1}{n-1} VSU^T USV^T \\ &= \frac{1}{n-1} VSSV^T \end{aligned}$$

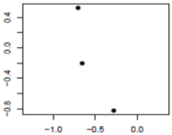
- The PC vectors are given by the columns of V
- The PC scores are given by UD, where $D = SS$

Another Way of Thinking about SVD

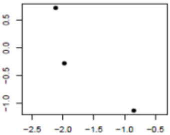
Observations in measurement space = (Observations in PC space) $\xleftarrow{\text{rotation}}$ Inverse of Eigenvectors

$$A = U\,S\,V^T$$

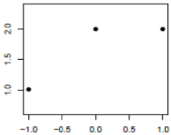
e.g. $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \\ -1 & 1 \end{bmatrix} =$



U



US (note change of scale)



USV^T

Applications of SVD

- Pseudoinverse of an arbirary matrix
- Matrix approximation
- Motivates the Biplot and Correspondence Analysis

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 A find matrix A^+ such that:

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

Moore-Penrose Inverse via SVD:

$$
\begin{aligned}
\text{if } A &= USV^T \\
A^+ &= VS^+U^T
\end{aligned}
$$

where S^+ has the reciprocal of non-zero elements of S.

SVD for Matrix Approximation

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

$$\tilde{A} = US^{\star}V^T$$

where:

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

Biplots

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

Given data matrix X, write

$$\underset{(n \times p)}{X} = \underset{(n \times p)}{U} \; \underset{(p \times p)}{S} \; \underset{(p \times p)}{V^T}$$

$$\tilde{X}_k = U \, S^* \, T \qquad \text{(}k\text{-dimensional approximation to } X\text{)}$$

reduce $\tilde{X}$ to a product

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

where $G = U(S^*)^{\alpha}$ (row effects) $\quad H^T = (S^*)^{1-\alpha} V^T$ (column effects)

if $\alpha = 1$, PCs are "sphered"

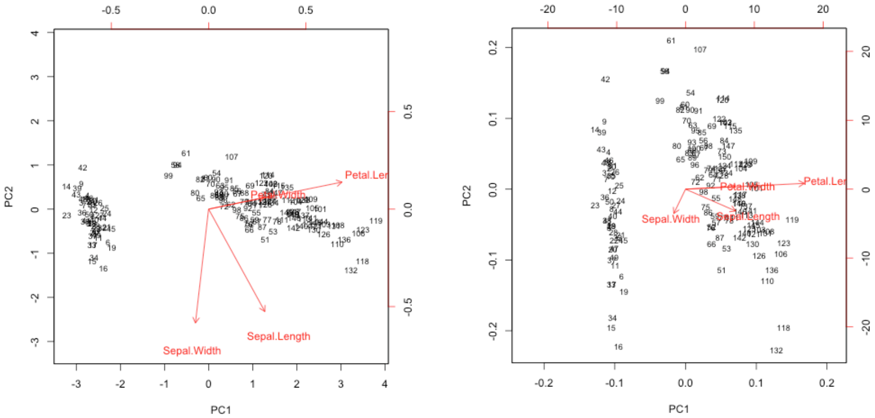
Biplots

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

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”)
- $\alpha = 0.5$, optimally approximates observational values (“symmetric biplot”)

Biplots, Example



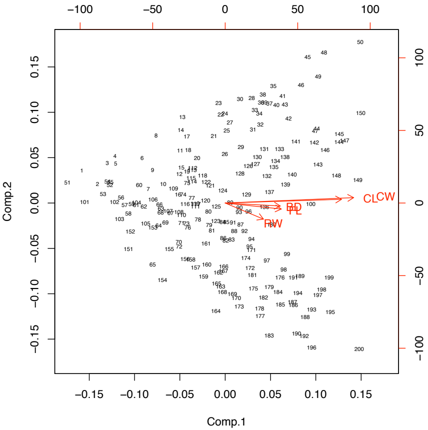
alpha = 0

alpha = 1

Anderson’s famous iris data set.

Biplots, Example II

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



Biplot ($\alpha = 1$) for dataset of five morphological measurements on crabs.