

Scientific Computing for Biologists

Singular Value Decomposition and Biplots

Instructor: Paul M. Magwene

18 October 2011

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
- SVD in Python
- Applications of SVD in R and Python
 - 'Seriation' using SVD
 - Matrix approximation and image compression using SVD

Eigendecomposition

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

where:

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

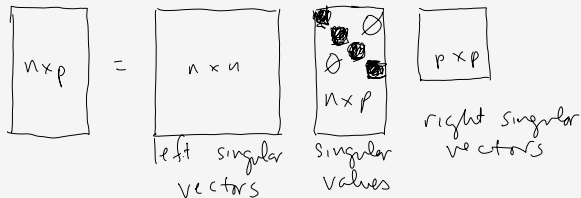
when A is real-valued and symmetric than U is orthogonal.

Singular Value Decomposition

$$A = U S V^T$$

$(n \times p)$ $(n \times n)$ $(n \times p)$ $(p \times p)$

assume $n \geq p$



when written like this U & V are orthonormal

• sometimes written as

$$A = U S V^T$$

$(n \times p)$ $(p \times p)$ $(p \times p)$

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:

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

- U and V are orthonormal matrices \leadsto Orthonormal matrices represent rigid rotations (or rotation plus reflection)
- Diagonal matrices represent “stretching”

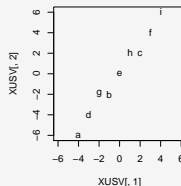
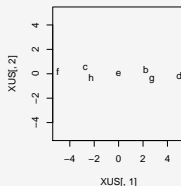
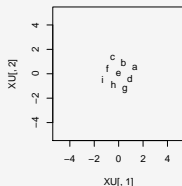
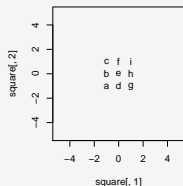
SVD Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = U S V^T$$

$$\text{where } U = \begin{bmatrix} -0.41 & -0.91 \\ -0.91 & 0.41 \end{bmatrix} \quad S = \begin{bmatrix} 5.47 & 0 \\ 0 & 0.37 \end{bmatrix}$$

$$V^T = \begin{bmatrix} -0.58 & -0.82 \\ 0.82 & -0.58 \end{bmatrix}$$

Geometry



Relationship of SVD to Eigendecomposition

$$A = U S V^T, \text{ let } D = S S$$

$$\begin{aligned} A^T A &= (V S U^T)(U S V^T) = V S \underbrace{U^T U}_= I S V^T \\ &= V D V^T \end{aligned}$$

so the singular values S_{ii} are $\sqrt{D_{ii}}$ where
 D_{ii} are the eigenvalues of $A^T A$

columns of V are the eigenvectors of $A^T A$

Using SVD to do PCA

let X be a mean-centered data matrix

covariance
matrix of X

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

By SVD we can write $X = U S V^T$

$$\begin{aligned} C &= \frac{1}{n} V S U^T U S V^T \\ &= \frac{1}{n} V S^2 V^T \end{aligned}$$

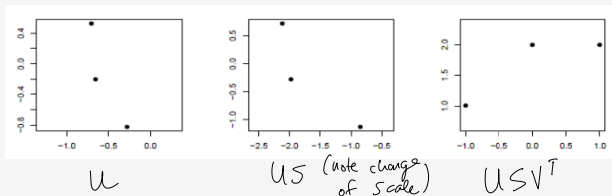
Principal Components given by columns of V

PC Scores given by $U D$

Another Way of Thinking about SVD

Observations
in measurement
space A = $\begin{pmatrix} \text{Observations} \\ \text{in PC} \\ \text{space} \end{pmatrix} \xleftarrow{\text{rotation}} \text{Inverse of} \\ U S V^T$

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



Applications of SVD

- Pseudoinverse of an arbitrary 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:

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

where:

$$S_{ii}^* = S_{ii} \text{ for } i \leq k$$

$$S_{ii}^* = 0 \text{ for } i > k$$

Biplots

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

Given data matrix X , write

$$X = U S V^T$$

$(n \times p)$ $(n \times p)$ $(p \times p)$ $(p \times p)$

$$\tilde{X}_k = U S^* T \quad \begin{matrix} k\text{-dimensional} \\ \text{(approximation to } X\text{)} \end{matrix}$$

reduce \tilde{X} to a product

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

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

(row effects) (column effects)

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

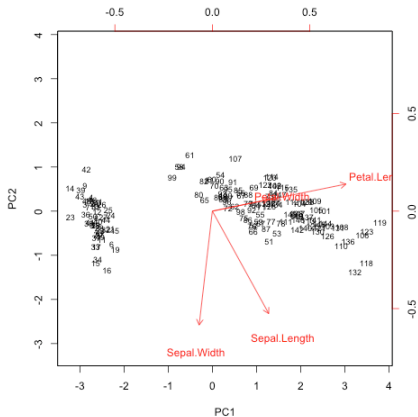
Biplots

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

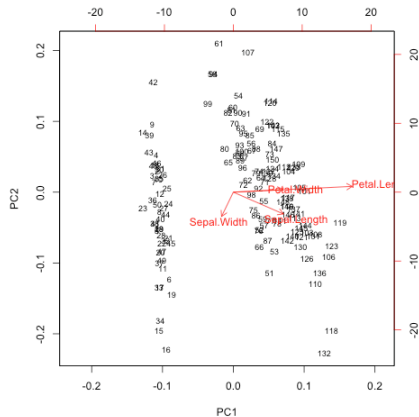
Different choices of α emphasize different relationships in the data.

- $\alpha = 0$, 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

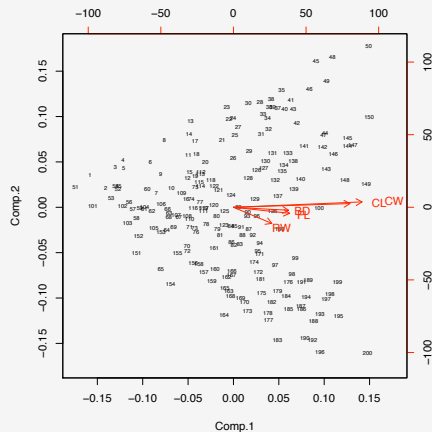


alpha = 1

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.