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

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

where:

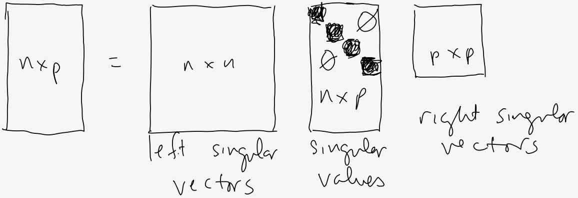
- **U** is a matrix of eigenvectors (in columns)
- **D** is a diagonal matrix with eigenvalues along diagonal.

when **A** is real-valued and symmetric than **U** is orthgonal.

Singular Value Decomposition

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

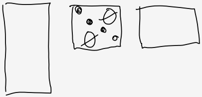
assume $n \geq p$



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 $\mathbf{A}$ is equivalent to the number of non-zero singular values in $\mathbf{A} = \mathbf{USV}^T$

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

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

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

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

Geometric Interpretation of SVD

Any matrix, $\mathbf{A}_{n \times p}$, 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:

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

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

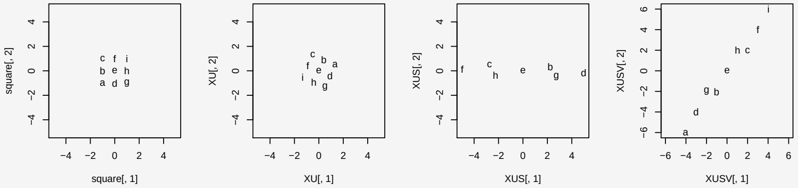
SVD Example

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

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

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

Geometry



Relationship of SVD to Eigendecomposition

$A = U S V^T$, let $D = SS$

$A^T A = (V S U^T)(U S V^T) = V S \underbrace{U^T U}_{= I \text{ (since orthonormal)}} S V^T$

$= V D V^T$

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 $X = USV^T$

$$C = \frac{1}{n} V S U^T U S V^T$$

$$= \frac{1}{n} V S^2 V^T$$

Principal Components given by columns of V

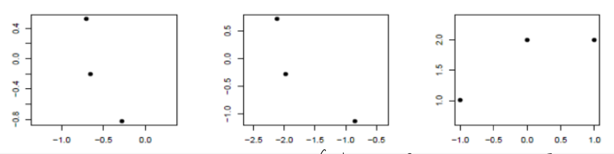
PC Scores given by UD

Another Way of Thinking about SVD

Observations in measurement space = (Observations in PC space) ← rotation — Inverse of Eigenvectors

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

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



U $\qquad$ US (note change of scale) $\qquad$ 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 $\mathbf{A}$ find matrix $\mathbf{A}^+$ such that:

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

Moore-Penrose Inverse via SVD:

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

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

SVD for Matrix Approximation

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

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

where:

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

Biplots

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

Given data matrix X, write

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

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

reduce $\tilde{X}$ to a product

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

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

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

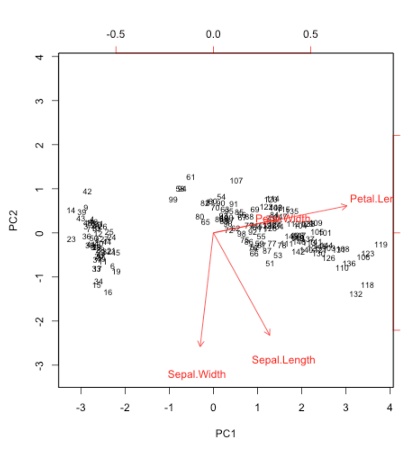
Biplots

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

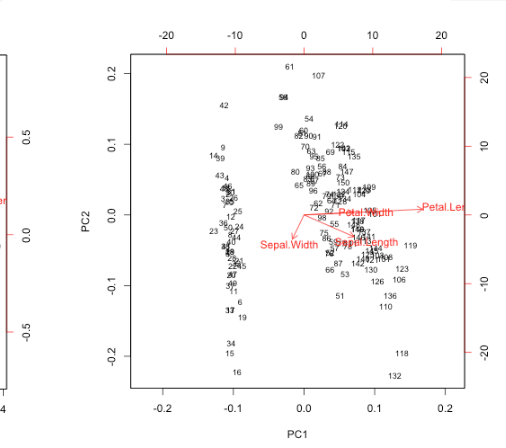
Different choices of α emphasize different relationships in the data.

- $\alpha = 0$, 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