Singular Value Decomposition (SVD)

Ariel Cintrón, Ph.D.

Reference

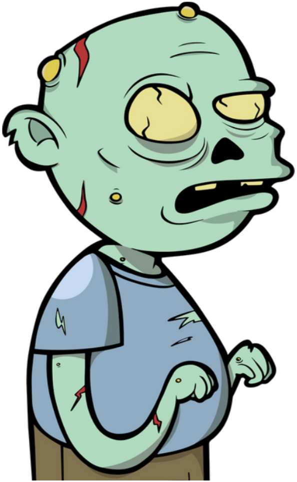
Chapter 10 Eigenvalues and Singular Values of C. Moler's *Numerical Computing with MATLAB*.

`http://www.mathworks.com/moler/chapters.html`

Zombie Math

Zombie Math ...Decompose

Zombie Math ...Decompose



Eigenvalue Decomposition

An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector x that satisfy

$$A\mathrm{x} = \lambda \mathrm{x}.$$

Eigenvalue Decomposition

An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector x that satisfy

$$A\mathrm{x} = \lambda \mathrm{x}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the n distinct eigenvalues of a square matrix A, let $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ be a set of corresponding eigenvectors.

Eigenvalue Decomposition

An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector x that satisfy

$$A\mathrm{x} = \lambda \mathrm{x}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the n distinct eigenvalues of a square matrix A, let $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ be a set of corresponding eigenvectors. Let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and let S denote the $n \times n$ matrix whose jth column is x_j.

Eigenvalue Decomposition

An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector x that satisfy

$$A\mathrm{x} = \lambda \mathrm{x}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the n distinct eigenvalues of a square matrix A, let $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ be a set of corresponding eigenvectors. Let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and let S denote the $n \times n$ matrix whose jth column is x_j. Then,

$$AS = S\Lambda,$$

which implies the *eigenvalue decomposition* of A:

Eigenvalue Decomposition

An *eigenvalue* and *eigenvector* of a square matrix A are a scalar λ and a nonzero vector x that satisfy

$$A\mathrm{x} = \lambda \mathrm{x}.$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the n distinct eigenvalues of a square matrix A, let $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ be a set of corresponding eigenvectors. Let $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and let S denote the $n \times n$ matrix whose jth column is x_j. Then,

$$AS = S\Lambda,$$

which implies the *eigenvalue decomposition* of A:

$$A = S\Lambda S^{-1}.$$

Singular Value Decomposition

What about if A is a rectangular matrix?

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition?

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition? Is there another decomposition or factorization?

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition? Is there another decomposition or factorization?

A **singular value** and pair of **singular vectors** of a rectangular matrix A are a nonnegative scalar σ and two nonzero vectors u and v such that

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition? Is there another decomposition or factorization?

A **singular value** and pair of **singular vectors** of a rectangular matrix A are a nonnegative scalar σ and two nonzero vectors u and v such that

$$A\mathsf{v} = \sigma \mathsf{u},$$

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition? Is there another decomposition or factorization?

A **singular value** and pair of **singular vectors** of a rectangular matrix A are a nonnegative scalar σ and two nonzero vectors u and v such that

$$A\mathsf{v} = \sigma \mathsf{u}, \qquad A^* \mathsf{u} = \sigma \mathsf{v},$$

Singular Value Decomposition

What about if A is a rectangular matrix? Is there an eigenvalue decomposition? Is there another decomposition or factorization?

A **singular value** and pair of **singular vectors** of a rectangular matrix A are a nonnegative scalar σ and two nonzero vectors u and v such that

$$A\mathrm{v} = \sigma \mathrm{u}, \qquad A^*\mathrm{u} = \sigma \mathrm{v},$$

where A^* denotes the *Hermitian transpose* or the complex conjugate transpose of A (we have been calling this operator the adjoint). In other words, $A^* = A^H = \overline{A^T} = \overline{A}^T$.

Singular Value Decomposition

Using matrix notation we can write the defining equations of of singular values and vectors as follows

$$AV = U\Sigma,$$
$$A^*U = V\Sigma^*,$$

Singular Value Decomposition

Using matrix notation we can write the defining equations of of singular values and vectors as follows

$$AV = U\Sigma,$$
$$A^*U = V\Sigma^*,$$

where the matrix Σ is basically diagonal, because all of its off diagonal entries are zero, but its main diagonal stores the σ_j's values where $\sigma_1 \geq \ldots \sigma_{n-1} \geq \sigma_n \geq 0$.

Singular Value Decomposition

Using matrix notation we can write the defining equations of of singular values and vectors as follows

$$AV = U\Sigma,$$
$$A^*U = V\Sigma^*,$$

where the matrix Σ is basically diagonal, because all of its off diagonal entries are zero, but its main diagonal stores the σ_j's values where $\sigma_1 \geq \ldots \sigma_{n-1} \geq \sigma_n \geq 0$.

In other words, the singular values appear in decreasing order, with σ_1 being the largest.

Singular Value Decomposition

Using matrix notation we can write the defining equations of of singular values and vectors as follows

$$AV = U\Sigma,$$
$$A^*U = V\Sigma^*,$$

where the matrix Σ is basically diagonal, because all of its off diagonal entries are zero, but its main diagonal stores the σ_j's values where $\sigma_1 \geq \ldots \sigma_{n-1} \geq \sigma_n \geq 0$.
In other words, the singular values appear in decreasing order, with σ_1 being the largest. The columns of matrices U and V store the normalized singular vectors, i.e., they satisfy

$$U^*U = I$$
$$V^*V = I,$$

Singular Value Decomposition

Using matrix notation we can write the defining equations of of singular values and vectors as follows

$$AV = U\Sigma,$$
$$A^*U = V\Sigma^*,$$

where the matrix Σ is basically diagonal, because all of its off diagonal entries are zero, but its main diagonal stores the σ_j's values where $\sigma_1 \geq \ldots \sigma_{n-1} \geq \sigma_n \geq 0$.
In other words, the singular values appear in decreasing order, with σ_1 being the largest. The columns of matrices U and V store the normalized singular vectors, i.e., they satisfy

$$U^*U = I$$
$$V^*V = I,$$

Therefore,

$$A = U\Sigma V^*$$

Example

Let $A = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$.

Example

Let $A = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$. Suppose $A = USV^T$, where

Example

Let $A = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$. Suppose $A = USV^T$, where

$$U = \begin{bmatrix} -0.2691 & -0.6798 & 0.6822 \\ 0.9620 & -0.1557 & 0.2243 \\ -0.0463 & 0.7167 & 0.6959 \end{bmatrix}$$

Example

Let $A = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$. Suppose $A = USV^T$, where

$$U = \begin{bmatrix} -0.2691 & -0.6798 & 0.6822 \\ 0.9620 & -0.1557 & 0.2243 \\ -0.0463 & 0.7167 & 0.6959 \end{bmatrix}$$

$$S = \begin{bmatrix} 817.7597 & 0 & 0 \\ 0 & 2.4750 & 0 \\ 0 & 0 & 0.0030 \end{bmatrix}$$

Example

Let $A = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$. Suppose $A = USV^T$, where

$$U = \begin{bmatrix} -0.2691 & -0.6798 & 0.6822 \\ 0.9620 & -0.1557 & 0.2243 \\ -0.0463 & 0.7167 & 0.6959 \end{bmatrix}$$

$$S = \begin{bmatrix} 817.7597 & 0 & 0 \\ 0 & 2.4750 & 0 \\ 0 & 0 & 0.0030 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.6823 & -0.6671 & 0.2990 \\ 0.2287 & -0.1937 & -0.9540 \\ 0.6944 & 0.7193 & 0.0204 \end{bmatrix}$$

Principal Components

- Principal component analysis approximates a general matrix by a sum of a few "simple" matrices.

Principal Components

- Principal component analysis approximates a general matrix by a sum of a few “simple” matrices.
- By simple we mean *rank one*, in other words all the rows are multiples of each other.

Principal Components

- Principal component analysis approximates a general matrix by a sum of a few "simple" matrices.
- By simple we mean *rank one*, in other words all the rows are multiples of each other.
- For example, consider

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Principal Components

- Principal component analysis approximates a general matrix by a sum of a few “simple” matrices.
- By simple we mean *rank one*, in other words all the rows are multiples of each other.
- For example, consider

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

- Clearly, the rank of A is one, i.e., $r = 1$.

Rank-One Matrices

In fact, *every rank one matrix has the special form* $A = uv^T$, in other words it can be written as the product of a column vector times a row vector.

Rank-One Matrices

In fact, *every rank one matrix has the special form* $A = uv^T$, in other words it can be written as the product of a column vector times a row vector.

For this example, we have $u = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix}$ and $v^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, i.e.,

Rank-One Matrices

In fact, *every rank one matrix has the special form* $A = uv^T$, in other words it can be written as the product of a column vector times a row vector.

For this example, we have $u = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix}$ and $v^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, i.e.,

$$A = uv^T$$

Rank-One Matrices

In fact, *every rank one matrix has the special form* $A = uv^T$, in other words it can be written as the product of a column vector times a row vector.

For this example, we have $u = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix}$ and $v^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, i.e.,

$$A = uv^T$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank-One Matrices

In fact, *every rank one matrix has the special form* $A = uv^T$, in other words it can be written as the product of a column vector times a row vector.

For this example, we have $u = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix}$ and $v^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, i.e.,

$$A = uv^T$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

Principal Components

- Suppose A is any real $m \times n$ matrix with

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

Principal Components

- Suppose A is any real $m \times n$ matrix with

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

- this SVD can be re-written as

$$A = E_1 + E_2 + \cdots + E_p,$$

where $p = \min(m, n)$.

Principal Components

- Suppose A is any real $m \times n$ matrix with

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

- this SVD can be re-written as

$$A = E_1 + E_2 + \cdots + E_p,$$

  where $p = \min(m, n)$.
- The component matrices E_k are rank one outer products (column vector times row vector):

$$E_k = \sigma_k \mathrm{u}_k \mathrm{v}_k^T.$$

Principal Components

- Suppose A is any real $m \times n$ matrix with

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

- this SVD can be re-written as

$$A = E_1 + E_2 + \cdots + E_p,$$

  where $p = \min(m, n)$.
- The component matrices E_k are rank one outer products (column vector times row vector):

$$E_k = \sigma_k \mathrm{u}_k \mathrm{v}_k^T.$$

- Each column of E_k is a multiple of u_k, the kth column of U, and each row is a multiple of v_k^T, the transpose of the kth column of V.

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$
$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$
$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$
- The norm of each component matrix is the corresponding singular value
$$\|E_k\| = \sigma_k$$

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$
$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$
- The norm of each component matrix is the corresponding singular value
$$\|E_k\| = \sigma_k$$
- The contribution each E_k makes to reproducing A is determined by the size of the singular value.

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$

$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$

- The norm of each component matrix is the corresponding singular value

$$\|E_k\| = \sigma_k$$

- The contribution each E_k makes to reproducing A is determined by the size of the singular value.
- If the sum

$$A = E_1 + E_2 + \cdots + E_p,$$

is truncated after $r < p$ terms, then

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$
$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$
- The norm of each component matrix is the corresponding singular value
$$\|E_k\| = \sigma_k$$
- The contribution each E_k makes to reproducing A is determined by the size of the singular value.
- If the sum
$$A = E_1 + E_2 + \cdots + E_p,$$
is truncated after $r < p$ terms, then
$$A_r = E_1 + E_2 + \cdots + E_r,$$

Principal Component Analysis (PCA)

- The component matrices are orthogonal in the sense that for $j \neq k$
$$E_j E_k^T = \sigma_j \sigma_k \mathrm{u}_j \mathrm{v}_j^T \mathrm{v}_k \mathrm{u}_k^T = 0_{m \times m}$$
- The norm of each component matrix is the corresponding singular value
$$\|E_k\| = \sigma_k$$
- The contribution each E_k makes to reproducing A is determined by the size of the singular value.
- If the sum
$$A = E_1 + E_2 + \cdots + E_p,$$
is truncated after $r < p$ terms, then
$$A_r = E_1 + E_2 + \cdots + E_r,$$
- the result is A_r, a rank r approximation to the original matrix A.

Principal Component Analysis (PCA)

- The matrix A_r is the closest rank r approximation to A.

Principal Component Analysis (PCA)

- The matrix A_r is the closest rank r approximation to A.
- The error in this approximation is

$$\|A - A_r\| = \sigma_{r+1}$$

Principal Component Analysis (PCA)

- The matrix A_r is the closest rank r approximation to A.
- The error in this approximation is

$$\|A - A_r\| = \sigma_{r+1}$$

- Since the singular values are in decreasing order, the accuracy of the approximation increases as the rank increases.

Principal Component Analysis (PCA)

- The matrix A_r is the closest rank r approximation to A.
- The error in this approximation is

$$\|A - A_r\| = \sigma_{r+1}$$

- Since the singular values are in decreasing order, the accuracy of the approximation increases as the rank increases.
- PCA is used in a wide range of fields, including statistics, earth sciences, archeology, and image processing.

PCA Example

Suppose we measure the height and weight of six subjects and store the data as columns of A with SVD denoted as $A = USV^T$.

PCA Example

Suppose we measure the height and weight of six subjects and store the data as columns of A with SVD denoted as $A = USV^T$.

Let $A = \begin{bmatrix} 47 & 15 \\ 93 & 35 \\ 53 & 15 \\ 45 & 10 \\ 67 & 27 \\ 42 & 10 \end{bmatrix}$, where

PCA Example

Suppose we measure the height and weight of six subjects and store the data as columns of A with SVD denoted as $A = USV^T$.

$$\text{Let } A = \begin{bmatrix} 47 & 15 \\ 93 & 35 \\ 53 & 15 \\ 45 & 10 \\ 67 & 27 \\ 42 & 10 \end{bmatrix}, \text{ where } \quad U = \begin{bmatrix} -0.3153 & 0.1056 \\ -0.6349 & -0.3656 \\ -0.3516 & 0.3259 \\ -0.2929 & 0.5722 \\ -0.4611 & -0.4562 \\ -0.2748 & 0.4620 \end{bmatrix},$$

PCA Example

Suppose we measure the height and weight of six subjects and store the data as columns of A with SVD denoted as $A = USV^T$.

Let $A = \begin{bmatrix} 47 & 15 \\ 93 & 35 \\ 53 & 15 \\ 45 & 10 \\ 67 & 27 \\ 42 & 10 \end{bmatrix}$, where $U = \begin{bmatrix} -0.3153 & 0.1056 \\ -0.6349 & -0.3656 \\ -0.3516 & 0.3259 \\ -0.2929 & 0.5722 \\ -0.4611 & -0.4562 \\ -0.2748 & 0.4620 \end{bmatrix}$,

$S = \begin{bmatrix} 156.4358 & 0 \\ 0 & 8.7658 \end{bmatrix}$, and $V = \begin{bmatrix} -0.9468 & 0.3219 \\ -0.3219 & -0.9468 \end{bmatrix}$.

PCA Example

Note that $\sigma_1 = 156.4358$ is much larger than $\sigma_2 = 8.7658$.

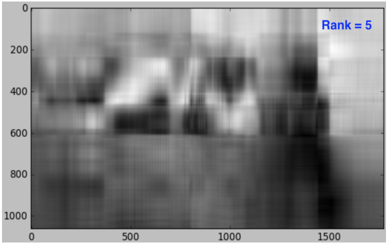
PCA Example

Note that $\sigma_1 = 156.4358$ is much larger than $\sigma_2 = 8.7658$.
A rank one approximation to A is

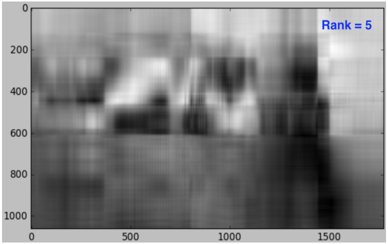
$$E_1 = \sigma_1 \mathrm{u}_1 \mathrm{v}_1^T = \begin{bmatrix} 46.7021 & 15.8762 \\ 94.0315 & 31.9657 \\ 52.0806 & 17.7046 \\ 43.3857 & 14.7488 \\ 68.2871 & 23.2139 \\ 40.6964 & 13.8346 \end{bmatrix}$$

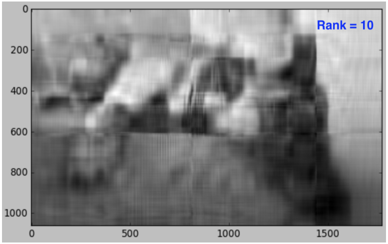
PCA Example in Digital Image Processing

PCA Example in Digital Image Processing



PCA Example in Digital Image Processing





PCA Example in Digital Image Processing