UNIT 6: The singular value decomposition.

María Barbero Liñán

Universidad Carlos III de Madrid
Bachelor in Statistics and Business

Mathematical methods II

2011-2012

- A **square** matrix is **symmetric** if $A^T = A$.
- A **square** matrix A is **orthogonally diagonalizable** if there exist an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1},$$

where P, D and A have all the same size.

Remember that...

an orthogonal matrix satisfies $P^T = P^{-1}$.

Theorems.

1. A $n \times n$ matrix A **orthogonally diagonalize if and only if** A **is a symmetric matrix**.
2. Let A be a symmetric matrix. **The eigenvectors of different eigenvalues are orthogonal.**

Steps to orthogonally diagonalize a matrix A.

1. Find the eigenvalues of A, which are the zeroes of the characteristic polynomial, $P_A(\lambda) = \det(A - \lambda I_n)$.
2. Find the eigenvectors of each eigenvalue.
    1. If an eigenvalue has more than one eigenvector, check if they are orthogonal.
    2. If they are not orthogonal, obtain an othogonal basis by Gram-Schmidt method.
3. Normalize those eigenvectors that are not unit.
4. The columns of P are the eigenvectors which determine a set of orthonormal vectors.
5. The diagonal matrix D has the eigenvalues in the main diagonal in the same order as the eigenvectors in P.

A VECTOR NORM is...

a function $\|\cdot\|: \mathbb{R} \to V$ that assigns a real-valued length to each vector in V and satisfies the following conditions:

1. $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$.
2. $\|x + y\| \leq \|x\| + \|y\|$.
3. $\|\lambda x\| = |\lambda|\|x\|$.

Examples of norms in $\mathbb{R}^n$.

1. **Euclidean norm or 2-norm:** $\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$ (the usual one).
2. **1-norm:** $\|x\|_1 = |x_1| + \cdots + |x_n|$.
3. **Infinity norm:** $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$.

The matrix norm of $A \in M_{m\times n}$ is defined as follows

$$\|A\|_{(m,n)} = \max_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \max_{\substack{x\in\mathbb{R}^n \\ \|x\|_{(n)}=1}} \|Ax\|_{(m)}.$$

Examples of matrix norms.

1. $$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{1\le k\le n} \left\{ \sum_{j=1}^{m} |a_{jk}| \right\}.$$
   It is the maximum of the 1-norm of each column of the matrix.
2. $$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1\le k\le m} \left\{ \sum_{j=1}^{n} |a_{kj}| \right\}.$$
   It is the maximum of the 1-norm of each row of the matrix.

The singular values σ of $A \in M_{m\times n}$

are the square roots of the eigenvalues λ of the square matrix $A^T A$, that is,

$$\sigma = \sqrt{\lambda},$$

where λ are the zeroes of the characteristic polynomial of $A^T A$, $P_{A^T A}(\lambda) = \det(A^T A - \lambda \mathrm{I}_n)$.

REMARK:

The matrix $A^T A$ is square and symmetric. It has size $n \times n$.

$\|A\|_2 = \max\{\sigma_1, \dots, \sigma_r\}$

where r is the number of the singular values of A.

The rank of A is equal to the number of nonzero singular values.

The image of the unit sphere...

under any matrix A of size $m \times n$ is an **hyperellipse**.

An hyperellipse is...

a surface obtained by stretching the unit sphere in $\mathbb{R}^n$ by some factors $\sigma_1, \ldots \sigma_r$ in some orthogonal directions.

The principal semiaxes of the hyperellipse are Av_i,

where v_i's are the eigenvectors of $A^T A$, and the semiaxes have length σ_i.

Singular value decomposition theorem (SVD).

Let A be a $m \times n$ matrix with rank r. Then there exists a unique $m \times n$ matrix, $\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$, for which the diagonal entries in the $r \times r$ matrix D are the r singular values of A, $\sigma_{\mathbf{1}} \geq \sigma_{\mathbf{2}} \geq \cdots \geq \sigma_{\mathbf{r}} > \mathbf{0}$; and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$\mathbf{A} = \mathbf{U \Sigma V^T}.$$

REMARK:

- The matrices U and V are not uniquely determined.
- The columns of U are called **left singular vectors of** A.
- The columns of V are called **right singular vectors of** A.

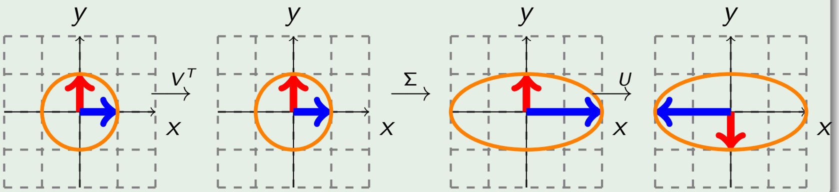
Steps to find the SVD of a $m \times n$ matrix A with $m \geq n$:

1. Find the eigenvalues λ of $A^T A$ and a set of ORTHONORMAL eigenvectors. (Use the Gram-Schmidt method to orthogonalize and normalize whenever is necessary).
2. Find Σ by writing the singular values $\sqrt{\lambda}$ in decreasing order in the main diagonal. The remaining entries of the matrix are zero up to obtain a $m \times n$ matrix, the same size as A.
3. Find V whose columns are the set of orthonormal eigenvectors written in the same order as the singular values in Σ. The size of V is $n \times n$.
4. Find U whose size is $m \times m$ and whose columns are:
    1. $u_i = \dfrac{1}{\sigma_i} A v_i$.
    2. Complete $\{u_1, \ldots, u_n\}$ to an orthonormal basis for $\mathbb{R}^m$ computing the orthogonal subspace of $\{u_1, \ldots, u_n\}$ and normalizing the spanning set of it.

Example of SVD of A.

$$A = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix} = \overset{\text{Rotate}}{\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}} \overset{\text{Stretch}}{\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}} \overset{\text{Rotate}}{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}} = U\Sigma V^T$$

The image of the unit sphere in $\mathbb{R}^2$ under A is an ellipse in $\mathbb{R}^2$ with center at $(0, 0)$ and semiaxis over the coordinate axis of length 2 and 1.



Reduced SVD

Let A be a $m \times n$ matrix with $m \geq n$ and SVD given by $U\Sigma V^T$.
The **reduced singular valued decomposition** of A is given by

$$
\begin{aligned}
A &= \begin{pmatrix} {\color{red}u_1} & \dots & {\color{red}u_n} & u_{n+1} & \dots & u_m \end{pmatrix} \begin{pmatrix} D_{n\times n} \\ 0_{(m-n)\times n} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_n \end{pmatrix} \\
&= \begin{pmatrix} {\color{red}u_1} & \dots & {\color{red}u_n} & u_{n+1} & \dots & u_m \end{pmatrix} \begin{pmatrix} D_{n\times n} V^T \\ 0_{(m-n)\times n} \end{pmatrix} \\
&= \begin{pmatrix} {\color{red}u_1} & \dots & {\color{red}u_n} \end{pmatrix} D_{n\times n} V^T.
\end{aligned}
$$

Remark: Note that to obtain the reduced singular value decomposition of A we do not need to extend the singular vectors in V to obtain a $m \times m$ matrix for U, we only use the left most $m \times n$ submatrix of U.

Moore-Penrose pseudoinverse $A^\dagger$ of A.

- If A is an invertible matrix, then $\mathbf{A}^\dagger = \mathbf{A}^{-1}$.
- If A is a $m \times n$ matrix with $m \geq n$ and SVD given by $U\Sigma V^T$, then **the Moore-Penrose pseudoinverse matrix is a** $n \times m$ **matrix** given by

$$\mathbf{A}^\dagger = (\mathbf{V}^\mathbf{T})^{-1}\mathbf{\Sigma}^\dagger\mathbf{U}^{-1} = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^\mathbf{T},$$

where

$$\Sigma^\dagger = \begin{pmatrix} D \\ 0 \end{pmatrix}^+ = \begin{pmatrix} D^{-1} & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & & \cdots & \frac{1}{\sigma_r} & 0 \end{pmatrix}.$$

Properties of the Moore-Penrose pseudoinverse:

1. $AA^{\dagger}A = A$.
2. $A^{\dagger}AA^{\dagger} = A^{\dagger}$.
3. $(AA^{\dagger})^T = AA^{\dagger}$.
4. $(A^{\dagger}A)^T = A^{\dagger}A$.

Application

When $m > n$, the Moore-Penrose pseudoinverse matrix gives a least-square solution of $Ax = b$.

- If A has full rank, $\mathbf{A}^{\dagger} = (\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^{\mathbf{T}}$.

  (Note: $A^{\dagger}b$ is a least-square solution of $Ax = b$ and it is exactly the solution to the normal equations in UNIT 5.)
- If A does not have full rank, we need SVD to obtain $A^{\dagger}$. A least-square solution of $Ax = b$ is given by $A^{\dagger}b$.

More details on the previous application when A does not have full rank

1. Find the SVD of A: $A = U\Sigma V^T$.
2. A least-square solution of $Ax = b$ is given by $A^\dagger b$ because

$$\begin{aligned}
(U\Sigma V^T)x = Ax &= b, \\
(U^T U)\Sigma V^T x &= U^T b, \\
\Sigma^\dagger \Sigma V^T x &= \Sigma^\dagger U^T b \quad \text{because } \Sigma^\dagger\Sigma = \mathrm{I}_{n\times n}, \\
(VV^T)x &= V\Sigma^\dagger U^T b, \\
\mathbf{x} &= \mathbf{V}\boldsymbol{\Sigma}^\dagger \mathbf{U}^\mathbf{T}\mathbf{b} = \mathbf{A}^\dagger \mathbf{b}.
\end{aligned}$$

Remark: Remember that the SVD is not unique because the matrices U and V are not unique. Hence, depending on the SVD we find different least-square solutions of $Ax = b$.