Table of Contents

Décomposition en valeurs singulières

Ce chapitre introduit une décomposition matricielle importante, la décomposition en valeurs singulières. Même si une telle décomposition n'est pas unique, chaque variante a des applications utiles. Comprendre ce chapitre

- Donner une motivation pour la décomposition en valeurs singulières d'une matrice.
- Définir les valeurs singulières et la décomposition en valeurs singulières.
- Calculer la décomposition en valeurs singulières d'une matrice.

Prérequis Les connaissances requises pour ce chapitre incluent les **valeurs propres** et **vecteurs propres** et l'**orthogonalité**.

Motivation

On considère la problématique d'identification de structures dans les données. A titre d'illustation, on s'intéresse en particulier à la recherche de structure dans les classements de films par des spectateurs :

Movie	Alice	Tom	Léa
Star Wars	5	4	1
Blade runner	5	5	0
Amélie	0	0	5
Delicatessen	1	0	4

Chaque spectateur a attribué une note entre 0 et 5 à chacun des filmls en fonction de leurs goûts. Que peut-on dire de ce classement?

Soit $\mathbf{A} \in \mathbb{R}^{n\times n}$. Si $\mathbf{A}$ est diagonalisable, alors $\mathbf{A} = \mathbf{PDP}^{-1}$ pour $\mathbf{P}, \mathbf{D} \in \mathbb{C}^{n\times n}$ telles que $\mathbf{P}$ soit inversible et $\mathbf{D}$ soit diagonale. Nous avons vu comment une telle décomposition permet de calculer aisément les puissances $\mathbf{A}^k$ pour tout entier positif k.

Dans le cas où $\mathbf{A} \in \mathbb{R}^{n\times n}$ est symétrique, nous avons vu que, $\mathbf{A} = \mathbf{QDQ}^T$ pour $\mathbf{Q}, \mathbf{D} \in \mathbb{R}^{n\times n}$ telles que $\mathbf{Q}$ soit orthogonale et $\mathbf{D}$ soit diagonale.

Definition

> Une matrice est dîte orthogonale si et seulement si $A^T A = I$ (ou A A^T =I). Autrement dit, $A^T = A^{-1}$, et les colonnes de A sont orthonormales deux à deux.

Soit $\vec{x} \in \mathbb{R}^n$. L'application $T : \mathbb{R}^n \to \mathbb{R}^n$ définie par $T(\vec{x}) = \mathbf{Q}\vec{x}$ est une application linéaire préservant les longueurs; on dit qu'elle est une **isométrie**. Or, si $S : \mathbb{R}^n \to \mathbb{R}^n$ est définie par $S(\vec{x}) = \mathbf{A}\vec{x}$, alors $S(x) = \mathbf{QDQ}^T\vec{x}$ commence par appliquer une isométrie, pour ensuite mettre à l'échelle les composantes résultantes, et finalement appliquer l'inverse de l'isométrie. Lorsque $n = 2$, on peut visualiser une telle suite de transformations comme une rotation/réflexion suivie d'une mise à l'échelle, suivie de la rotation/réflexion inverse.

Que se passer-t-il si nous remplaçons la dernière transformation par une autre isométrie? Autrement dit, quelles matrices peut-on décomposer sous la forme

$$\mathbf{PDQ}^T,$$

où $\mathbf{P}$ et $\mathbf{Q}$ sont orthogonales et $\mathbf{D}$ est diagonale? Il s'avère que toutes les matrices de $\mathbb{R}^{n\times n}$ ont une telle décomposition!

Théorème

> Pour chaque $\mathbf{A} \in \mathbb{R}^{n\times n}$, il existe une matrice diagonale $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ et des matrices orthogonales $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times n}$ telles que
>
> $$\mathbf{A} = \mathbf{U\Sigma V}^T.$$

Si $\mathbf{A} = \mathbf{U\Sigma V}^T$, où $\mathbf{U}$ et $\mathbf{V}$ sont orthogonales et $\mathbf{\Sigma}$ diagonale, alors la décomposition en valeurs singulières est une diagonalisation orthogonale de la matrice symétrique $\mathbf{A}^T\mathbf{A}$.

Si on choisit une matrice $\mathbf{A} \in \mathbb{R}^{n\times n}$ arbitrairement, peut-on utiliser une diagonalisation orthogonale de $\mathbf{A}^T\mathbf{A}$ pour obtenir $\mathbf{\Sigma}$ et $\mathbf{V}$?

Les valeurs propres de $\mathbf{A}^T\mathbf{A}$ sont réelles et non négatives.

En effet, on considère $B \in \mathbb{R}^n$ une matrice symétrique. Soit v un vecteur propre de B associé à la valeur propre λ. On a que

$$\begin{aligned}(Bv)^T v &= v^T B^T v \\ &= v^T Bv \\ &= \lambda v^T v\end{aligned}$$

Par définition, nous savons que $(Bv)^T v \geq 0$ (il s'agit de la norme induite par B) et $v^T v > 0$ (c'est la norme 2 de v). On en déduit que $\lambda \geq 0$.

Si on pose $B = A^T A$, nous avons bien la propriété annoncée au départ.

Construction de la SVD

On considère la diagonalisation de $\mathbf{A}^T\mathbf{A}$ donnée par $\mathbf{QDQ}^T$, c'est-à-d ire

$$A^T A = Q \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n^2 \end{bmatrix} Q^T.$$

On suppose ensuite que la décomposition en valeurs singulières de A existe. On peut donc écrire A sous la forme $A = U\Sigma V^T$, avec U et V des matrices orthogonales, ce qui conduite à

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T (U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V\Sigma^T \Sigma V^T \\ &= V \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix} V^T. \end{aligned}$$

On peut ainsi identifier les éléments des deux décompositions, nous obtenons $\sigma_k^2 = \lambda^k$ pour $1 \leq k \leq n$ et $\mathbf{V} = \mathbf{Q}$.

Nous devons maintenant déterminer la matrice orthogonale $\mathbf{U}$ telle que

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

Si toutes les valeurs propres soient positives alors $\mathbf{\Sigma}$ est inversible et

$$\mathbf{U} = \mathbf{AV\Sigma}^{-1}.$$

Nous devons vérifier que $\mathbf{U}$ est orthogonale :

$$\begin{aligned} U^T U &= (AV\Sigma^{-1})^T AV\Sigma^{-1}) \\ &= \Sigma^{-1} V^T A^T AV\Sigma^{-1} \\ &= \Sigma^{-1} V^T V D V^T V\Sigma^{-1} \\ &= \Sigma^{-1} D\Sigma^{-1} \\ &= I. \end{aligned}$$

$\mathbf{U}$ est donc orthogonale.

On en conclut que si la décomposition en valeurs singulières existe, alors nous pouvons la calculer simplement.

Si, au contraire, au moins une des valeurs propres est nulle, la situation est plus compliquée, mais conceptuellement semblable au cas précédent.

Valeurs Propres

Soit $\mathbf{A} \in \mathbb{R}^{m \times n}$. On a déjà vu que $\mathbf{AA}^T$ et $\mathbf{A}^T\mathbf{A}$ sont des matrices réelles et symétriques, et donc leurs valeurs propres sont des nombres réels. De plus, elles sont diagonalisable en base orthonormée.

Nous allons étudier quelques propriétés de ces valeurs propres.

1. les valeurs propres de $\mathbf{A}^T\mathbf{A}$ et les valeurs propres de $\mathbf{AA}^T$ sont non négatives.
2. Soit λ une valeur propre de $\mathbf{A}^T\mathbf{A}$ et v un vecteur propre non nul associé. On pose $u = Av$. On a donc que

$$\begin{aligned} AA^Tu &= AA^T(Av) \\ &= A(A^TA)v \\ &= A(\lambda v) \\ &= \lambda Av \\ &= \lambda u \end{aligned}$$

On en déduit ainsi que si v est vecteur propre associé à la valeur propre λ de A^TA, alors u est vecteur propre associé à la valeur propre λ de AA^T. Les matrices AA^T et A^TA possèdent les mêles valeurs propres.

De même, on peut montrer de la même façon que toute valeur propre de $\mathbf{AA}^T$ est aussi une valeur propre de $\mathbf{A}^T\mathbf{A}$. Par conséquent, $\mathbf{AA}^T$ et $\mathbf{A}^T\mathbf{A}$ ont les mêmes valeurs propres et elles ont les mêmes multiplicités (algébriques et géométriques).

Ce résultat se synthétise par la proposition suivante

Proposition

> Soit $\{v^{(1)}, \ldots, v^{(k)}\}$ une base de l'espace vectoriel des vecteurs propres de $\mathbf{A}^T\mathbf{A}$ associé à la valeur propre positive λ. Alors, les vecteurs $u^{(1)} = \mathbf{A}v^{(1)}, \ldots, u^{(k)} = \mathbf{A}v^{(k)}$ sont linéairement indépendants.

Démonstration

> Supposons que les vecteurs $u^{(1)}, \ldots, u^{(k)}$ ne soient pas linéairement indépendants. Alors il existe des scalaires $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ non tous nuls tels que
>
> $$\alpha_1 u^{(1)} + \cdots + \alpha_k u^{(k)} = 0.$$
>
> En multipliant chacun des côtés à gauche par $\mathbf{A}^T$, nous obtenons
>
> $$\alpha_1 \mathbf{A}^T u^{(1)} + \cdots + \alpha_k \mathbf{A}^T u^{(k)} = 0,$$
>
> ou, de façon équivalente,
>
> $$(\alpha_1\lambda)v^{(1)} + \cdots + (\alpha_k\lambda)v^{(k)} = 0.$$
>
> Puisque $\lambda > 0$ et que $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ sont non tous nuls, on en déduit que $\alpha_1\lambda, \ldots, \alpha_k\lambda$ sont non tous nuls. Ceci contredit l'indépendence linéaire de $v^{(1)}, \ldots, v^{(k)}$.

Soient $\lambda_1, \ldots, \lambda_r$ les valeurs propres positives de $\mathbf{A}^T\mathbf{A}$ ou $\mathbf{AA}^T$, ordonnées selon $\lambda_1 \geq \cdots \geq \lambda_r$. On appelle **valeurs singulières de $\mathbf{A}$** les quantités $\sigma_i = \sqrt{\lambda_i}$, $i = 1, \ldots, r$.

Cas général

Nous allons maintenant nous intéressé au cas des matrices rectangulaires.

Théorème

> Pour toute matrice $\mathbf{A} \in \mathbb{R}^{m\times n}$, il existe une matrice orthogonale $\mathbf{U} \in \mathbb{R}^{m\times m}$, une matrice orthogonale $\mathbf{V} \in \mathbb{R}^{n\times n}$, et une matrice diagonale $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$

> dont les éléments sont non négatifs, telles que
>
> $$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{1}$$

On dit qu'une matrice de dimension $m \times n$ est **diagonale** si tous ses éléments (i, j) sont nuls lorsque $i \neq j$.

Soit $\mathbf{V}\mathbf{D}\mathbf{V}^T$ une diagonalisation en base orthonormée de $\mathbf{A}^T\mathbf{A}$ telle que tous les éléments positifs de $\mathbf{D}$ sont placés dans les r premiers éléments de la diagonale.

Supposons aussi que les valeurs propres sont énumérées par $\lambda_1, \ldots, \lambda_r$, où $\lambda_1 \geq \cdots \geq \lambda_r$.

Pour $i = 1, \ldots r$, on dénote $\sqrt{\lambda_i}$ par σ_i, la i-ième colonne de $\mathbf{V}$ par $v^{(i)}$, et $\frac{1}{\sigma_i}\mathbf{A}v^{(i)}$ par $u^{(i)}$.

Notons que $u^{(i)}$ est unitaire pour chaque $i = 1, \ldots, r$.

De plus, lorsque $i \neq j$. Ainsi, on peut construire une base orthonormée $\{u^{(1)}, \ldots, u^{(m)}\}$ de $\mathbb{R}^m$, à partir de $\{u^{(1)}, \ldots, u^{(r)}\}$.

Soit $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ la matrice diagonale telle que le i-ième élément de la diagonale est σ_i pour $i = 1, \ldots, r$ et tous les autres éléments sont nuls. Alors,

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

Puisque $\mathbf{V}$ est orthogonale, on multiplie à droite de chaque côté par $\mathbf{V}^T$ afin d’obtenir

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

ce qui complète la démonstration du théorème.

Exemple

Soit $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. Nous obtenons une décomposition en valeurs singulières de la matrice $\mathbf{A}$.

Premièrement, les valeurs propres de $\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ sont obtenues en trouvant les racines du polynôme charactéristique \begin{align*}

$$\begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 2-\lambda & 1 \\ 0 & 1 & 1-\lambda \end{vmatrix}$$

= (1-\lambda)

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix}$$

- $$\begin{vmatrix} 1 & 1 \\ 0 & 1-\lambda \end{vmatrix}$$

= -\lambda(\lambda - 3)(\lambda -1) \end{align*} Ainsi, les valeurs propres sont 3, 1 et 0. Posons $\sigma_1 = \sqrt{3}$ et $\sigma_2 = 1$.

La forme échelonnée réduite de $\mathbf{A}^T\mathbf{A} - 3\mathbf{I}$ est $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$. Ainsi, une base orthonormée de $null(\mathbf{A}^T\mathbf{A} - 3\mathbf{I})$ est donnée par le vecteur $v^{(1)} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}$. Posons $u^{(1)} = \frac{1}{\sigma_1}\mathbf{A}v^{(1)} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

La forme échelonnée réduite de $\mathbf{A}^T\mathbf{A} - \mathbf{I}$ est $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Ainsi, une base orthonormée de $null(\mathbf{A}^T\mathbf{A} - \mathbf{I})$ est donnée par le vecteur $v^{(2)} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. Posons $u^{(2)} = \frac{1}{\sigma_2}\mathbf{A}v^{(2)} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

La forme échelonnée réduite de $\mathbf{A}^T\mathbf{A} - 0\mathbf{I}$ est $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Ainsi, une base orthonormée de $null(\mathbf{A}^T\mathbf{A} - 0\mathbf{I})$ est donnée par le vecteur $v^{(3)} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$.

En posant $\mathbf{U} = \begin{bmatrix} u^{(1)} & u^{(2)} \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$, et $\mathbf{V} = \begin{bmatrix} v^{(1)} & v^{(2)} & v^{(3)} \end{bmatrix}$, on obtient $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, où $\mathbf{U}$ et $\mathbf{V}$ sont des matrices orthogonales.

Les calculs permettant d'obtenir une décomposition en valeurs singulières sont très long à faire à la main. En pratique, nous utiliser des bibliothèques dédiées pour réaliser cette décomposition.