Décomposition en valeurs singulières (SVD)

1 Définitions

Les valeurs singulières d'une matrice A (pas forcément carrée) permettront d'obtenir des informations très utiles sur A (distance à la singularité, rang...) et sont utilisées dans le cadre de nombreux problèmes d'algèbre linéaire (systèmes linéaires, moindres carrés, régularisation...) avec des applications dans des domaines variés (traitement du signal, recherche d'informations, reconnaissance de formes...). Dans ce qui suit nous considérons le cas réel mais les résultats s'appliquent également au cas complexe.

Décomposition en valeurs singulières: Soit $A \in \mathbb{R}^{m\times n}$. Il existe $U \in \mathbb{R}^{m\times m}$ et $V \in \mathbb{R}^{n\times n}$ matrices orthogonales telles que:

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T, \text{ si } m \geq n$$

$$A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V^T, \text{ si } m \leq n$$

où $\Sigma = diag(\sigma_1, \ldots, \sigma_p)$ avec $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ et $p = min(m,n)$.

Cette décomposition de A est appelée décomposition en valeurs singulières (singular value decomposition - SVD). Les scalaires σ_i sont appelés les valeurs singulières de A et sont uniques.
Les colonnes de U (vecteurs notés $u_1, \ldots, u_m$) et de V (vecteurs notés $v_1, \ldots, v_n$) sont appelées respectivement les vecteurs singuliers à gauche et à droite de A. Ils sont uniques au signe près.
Le rang de A correspond alors au nombre de valeurs singulières non nulles.

Python: np.linalg.svd(A)

Remarque: Si $A \in \mathbb{C}^{m\times n}$, les matrices U et V de la SVD de A sont unitaires ($U^*U = I$).

SVD réduite: Dans le cas où $m > n$ la SVD peut s'écrire $A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ où U_1 est $m \times n$ et Σ et V sont $n \times n$, ce qui donne

$$A = U_1 \, \Sigma \, V^T.$$

On parle alors de SVD réduite (ou "thin" SVD) que l'on utilise souvent en pratique. En effet dans le cas $m > n$, U_2 est inutile puisque multipliée par un blocs de valeurs nulles.

De même, lorsque $m < n$, on a $A = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}$, et la SVD réduite s'écrit alors

$$A = U \, \Sigma \, V_1^T.$$

Dans le cas où $m = n$, la SVD coïncide avec la SVD réduite.

Python: np.linalg.svd(A, full_matrices=False)

2 Propriétés

On suppose que l'on a la SVD réduite $A = U\ \Sigma\ V^T$ avec $rang(A) = r \leq p = min(m,n)$, $\Sigma = diag(\sigma_1, \ldots, \sigma_p)$ et $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$.

SVD développée: avec les notations précédentes, on a

$$\begin{aligned} A &= U\ \Sigma\ V^T \\ &= \left[\begin{array}{ccc} u_1 & \ldots & u_p \end{array}\right] \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{pmatrix} \left[\begin{array}{ccc} v_1 & \ldots & v_p \end{array}\right]^T \\ &= \left[\begin{array}{ccc} u_1 & \ldots & u_p \end{array}\right] \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{pmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_p^T \end{bmatrix} \\ &= \left[\begin{array}{ccc} u_1 & \ldots & u_p \end{array}\right] \begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_p v_p^T \end{bmatrix} \end{aligned}$$

Comme $\forall k > r,\ \sigma_k = 0$, on obtient donc

$$\boxed{A\ =\ \sum_{i=1}^{r} \sigma_i\ u_i\ v_i^T}$$

Cela montre que toute matrice de rang r peut s'écrire comme combinaison linéaire de r matrices $u_i v_i^T$ de rang 1.

Relations entre valeurs singulières et vecteurs singuliers: Exprimons la SVD en faisant apparaitre les vecteurs u_i et v_i.

$$\begin{aligned} A &= U\ \Sigma\ V^T \\ A\ V &= U\ \Sigma\ \ \text{(car V est orthogonale)} \end{aligned}$$

D'où:

$$\boxed{\forall i,\ A\ v_i = \sigma_i\ u_i}$$

De la même manière, en utilisant le fait que $A^T = V\ \Sigma\ U^T$ on obtient

$$\boxed{\forall i,\ A^T\ u_i = \sigma_i\ v_i}$$

Ainsi les σ_i nous informent de la distorsion qui apparait sous l'effet de A pour les vecteurs orthonormés v_i (respectivement sous l'effet de A^T pour les vecteurs orthonormés u_i).

Remarques:

- De $A = U\Sigma V^T$ et $A^T = V\Sigma U^T$ on déduit que A et A^T ont les mêmes valeurs singulières.
- Si $A = U\ \Sigma\ V^T$ est une SVD de A inversible alors $\boxed{A^{-1} = V\ \Sigma^{-1}\ U^T}$ est une SVD de A^{-1}.

- La SVD peut être vue comme une généralisation de la décomposition spectrale d'une matrice symétrique $A = U\ \Sigma\ U^T$ mais contrairement à la décomposition spectrale, la SVD existe pour toute matrice, y compris rectangulaire. Si A est symétrique, on a donc $\forall i, \sigma_i = |\lambda_i|$.

Liens avec les valeurs/vecteurs propres de A^TA (à démontrer):

- Les σ_i^2 sont les valeurs propres de A^TA (ou de AA^T).
- Les v_i sont des vecteurs propres de A^TA.
- Les u_i sont des vecteurs propres de AA^T.

3 Interprétation géométrique

Prenons le cas où $A \in \mathbb{R}^{n\times n}$ est inversible (et donc $rang(A) = n$). On note $S = \{x \in \mathbb{R}^n \ / \ \|x\|_2 = 1\}$ (sphère unité de $\mathbb{R}^n$). On montre ci-dessous que **l'image de S par A (noté $A.S$) est un ellipsoïde dont les demi-axes ont pour longueur les valeurs singulières de A et chaque v_k a pour image par A le k-ème demi-axe de $A.S$.**

Soit $y \in A.S$, alors $\exists x \in S,\ y = Ax$ et l'on a:

$$\begin{aligned}
\|x\|_2^2 &= \left\|A^{-1}y\right\|_2^2 \\
&= \left\|V\Sigma^{-1}U^Ty\right\|_2^2 \\
&= \left\|\Sigma^{-1}U^Ty\right\|_2^2 \\
&= \left\|\Sigma^{-1}z\right\|_2^2 \text{ en posant } z = U^Ty = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \\
&= \sum_{i=1}^{n} \frac{z_i^2}{\sigma_i^2}.
\end{aligned}$$

Comme $x \in S$ on a donc

$$1 = \sum_{i=1}^{n} \frac{z_i^2}{\sigma_i^2} \text{ avec } z \in U^TA.S \tag{1}$$

L'équation ci-dessus montre que $U^TA.S$ est un ellipsoïde ayant des demi-axes de longueur σ_i. Il en est de même pour $A.S = U(U^TA.S)$ car U est orthogonale et donc conserve les distances.
D'après l'équation (1), $U^TA.S$ a des axes selon la base canonique $e_1, \ldots, e_n$.
Par définition de U, on a $\forall k, Ue_k = u_k$ donc $A.S = U(U^TA.S)$ aura des axes dirigés selon les colonnes de U, c'est à dire les vecteurs singuliers à gauche de A.
De $AV = U\Sigma$ on déduit $\forall k, Av_k = \sigma_k u_k$ i.e que chaque v_k est un point de S dont l'image par A est le k-ème demi-axe de $A.S$.

4 Conditionnement d'une matrice

Definition 1. *Le conditionnement (en norme 2) d'une matrice inversible $A \in \mathbb{R}^{n\times n}$ est défini par $\kappa_2 = \|A\|_2 \left\|A^{-1}\right\|_2$.*

Python: np.linalg.cond(A)

On montre que (voir exercice) $\|A\|_2 = \sigma_1$ et $\|A^{-1}\|_2 = 1/\sigma_n$. Par conséquent on a $\boxed{\kappa_2 = \dfrac{\sigma_1}{\sigma_n}}$ qui, pour reprendre l'interprétation géométrique précédente mesure le degré de distorsion de la sphère unité sous l'action de A.

On observe que $\kappa_2 \geq 1$ et que, si $\kappa_2 = 1$, alors A est un multiple d'une matrice orthogonale.

Systèmes linéaires et SVD: De la même manière que pour la SVD développée de A, on a pour A^{-1}

$$\begin{aligned} A^{-1} &= V\Sigma^{-1}U^T \\ &= \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1}u_1^T \\ \vdots \\ \frac{1}{\sigma_n}u_n^T \end{bmatrix} \\ &= \sum_{i=1}^{n} \frac{1}{\sigma_i} v_i u_i^T \end{aligned}$$

La solution du système $Ax = b$ s'écrit donc

$$\begin{aligned} x &= A^{-1}\, b \\ &= \sum_{i=1}^{n} \frac{1}{\sigma_i} v_i \underbrace{u_i^T b}_{\in \mathbb{R}} \end{aligned}$$

D'où

$$x = \sum_{i=1}^{n} \left(\frac{u_i^T b}{\sigma_i} \right) v_i \tag{2}$$

Si la matrice est mal conditionnée, i.e si elle a des valeurs singulières petites par rapport à la plus grande, alors la division par ces petites valeurs dans l'équation (2) va amplifier les perturbations numériques sur les données (en particulier b).
On peut alors régulariser la solution en ne conservant que les termes pour lesquels $\sigma_i \geq \tau$ où τ est un seuil en dessous duquel les erreurs deviennent trop importantes. La solution approchée peut alors s'exprimer

$$\tilde{x} = \sum_{\sigma_i \geq \tau} \left(\frac{u_i^T b}{\sigma_i} \right) v_i.$$

On parle alors de SVD tronquée (truncated SVD ou TSVD).

5 Exercice: valeurs singulières extrêmes

Soit $A \in \mathbb{R}^{n \times n}$ telle que $rg(A) = n$ et $A = U\Sigma V^T$ sa décomposition en valeurs singulières avec $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$.

Montrer que $\|\Sigma\|_2 = \sigma_1$.

De la même manière montrer que $\|\Sigma^{-1}\|_2 = \frac{1}{\sigma_n}$.

Vérifier que l'on a $\kappa_2 = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$.