Décomposition en valeurs singulières (SVD)

Fouille de données avancée (2016-2017)

UFR MIME

Université Lille 3

30 Novembre 2016

Sommaire

Pourquoi SVD

Décomposition en valeur singulière ou SVD (Singular Value Decomposition) est motiver par deux opérations souvent fait dans l'analyse des données :

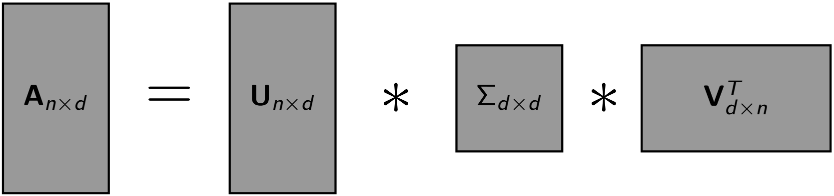
- **Découplage** : Séparation dans les composantes indépendantes pour faciliter analyse
- **Triage** : Ordonnancements de contributions par leur importance ou capacité d'explication

Parmi 100s des décompositions, le SVD reste une transformation puissante

Décomposition SVD

$\mathbf{A} \in \mathbb{R}^{n \times d}$ est une matrice, il existe une décomposition

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



- $\Sigma \in \mathbb{R}^{d \times d}$ est diagonale avec les entrées positives $\sigma_i > 0$
- $\mathbf{U} \in \mathbb{R}^{n \times d}$ vecteurs singuliers gauches (colonnes orthogonales)
- $\mathbf{V} \in \mathbb{R}^{d \times d}$ vecteurs singuliers droits (colonnes et rangs orthogonales)
- $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ et $\mathbf{U}\mathbf{U}^T = \mathbf{I}_d$ et $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, ..., \sigma_d)$ ou les valeurs singulières sont ordonnes : $\sigma_1 > \sigma_2, ..., > \sigma_d$

Produit extérieur

$$A_{n,d} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,d} \end{bmatrix} = U\Sigma V^T = \sum_{i=1}^{d} \sigma_i u_i v_i$$

$$A_{n,d} = \begin{bmatrix} | & | & | & | \\ | & | & | & | \\ u_1 & u_2 & \dots & u_d \\ | & | & | & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{bmatrix} \begin{bmatrix} ---v_1^T--- \\ ---v_2^T--- \\ \vdots \\ ---v_d^T--- \end{bmatrix}$$

$$\boxed{\mathbf{A}} = \begin{matrix} \sigma_1 \times -v_1^T- \\ | \\ u_1 \\ | \end{matrix} + \begin{matrix} \sigma_2 \times -v_2^T- \\ | \\ u_2 \\ | \end{matrix} + \dots \begin{matrix} \sigma_n \times -v_n^T- \\ | \\ u_n \\ | \end{matrix}$$

Produit scalaire

Étant donnée les matrices $\mathbf{U}, \Sigma, \mathbf{V}$, et u_i une ligne de $\mathbf{U}$ et v_j est une colonne de $\mathbf{V}^T$, $u_i, v_j \in \mathbb{R}^d$, on peut écrire :

$$M_{ij} = \sigma_k(u_i \cdot v_j) = \sum_k u_{ik}\sigma_k v_{jk} \tag{1}$$

$$M_{n,n} = \begin{bmatrix} ---u_1--- \\ \vdots \\ ---u_i--- \\ \vdots \\ ---u_n--- \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{bmatrix} \begin{bmatrix} | & | & | & | & | \\ | & | & | & | & | \\ v_1 & \ldots & v_j & \ldots & v_n \\ | & | & | & | & | \end{bmatrix}$$

- M serve comme une mesure de similarité entres des vecteurs u_i, v_j,
- Assez utile pour les systèmes de recommandations.

Remarque

Les valeur singulières (non-nul) de $\mathbf{A}$ sont les racines carrées de valeurs propres (non-nul) de $\mathbf{A}^T\mathbf{A}$ ou $\mathbf{A}\mathbf{A}^T$

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A} &= (\mathbf{U}\Sigma\mathbf{V}^T)^T(\mathbf{U}\Sigma\mathbf{V}^T) \\
&= (\mathbf{V}^{T^T}\Sigma^T\mathbf{U}^T)(\mathbf{U}\Sigma\mathbf{V}^T) \\
&= \mathbf{V}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T \\
&= \mathbf{V}\Sigma\mathbf{I}\Sigma\mathbf{V}^T \\
&= \mathbf{V}\Sigma^2\mathbf{V}^T = \mathbf{X}\Lambda\mathbf{X}^T \\
\mathbf{A}\mathbf{A}^T &= (\mathbf{U}\Sigma\mathbf{V}^T)(\mathbf{U}\Sigma\mathbf{V}^T)^T = \mathbf{U}\Sigma^2\mathbf{U}
\end{aligned} \tag{2}
$$

Comment calculer le SVD

Le nombre de valeurs positives singulières est égal au rang de la matrice.

- Donnée $\mathbf{X}_1 = \mathbf{A}\mathbf{A}^T$ et $\mathbf{X}_2 = \mathbf{A}^T\mathbf{A}$.
- Décomposition des matrices $\mathbf{X}_1 = \mathbf{Q}_1 \Lambda \mathbf{Q}_1^{-1}$ et $\mathbf{X}_2 = \mathbf{Q}_2 \Lambda \mathbf{Q}_2^{-1}$ en éléments propres.
- Par contre on sais $\mathbf{A}\mathbf{A}^T = \mathbf{U}\Sigma^2\mathbf{U}$ et $\mathbf{A}^T\mathbf{A} = \mathbf{V}\Sigma^2\mathbf{V}^T$
- Calculer la racine carrée des valeur propres positif $\sigma_i = \sqrt{\lambda_i}$
- Finalement rendre $\mathbf{U}, \mathbf{V}, \Sigma$ vu $\mathbf{Q}_1, \mathbf{Q}_2, \Lambda$

Calcul : En utilisant les algorithmes de calcule des vecteurs propres, on peut calculer la décomposition SVD.

Question : quelle est la différence entre la décomposition en valeur propres et la décomposition SVD ?

Troncation des valeurs singulières

- $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots \sigma_n u_n v_n^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$

Troncation des valeurs singulières

- $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots \sigma_n u_n v_n^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$
- On peut approximer $\mathbf{A}$ en prenant seulement les premières r-colonnes de $\mathbf{U}$ (et $\mathbf{V}$) et les premières r-valeurs singulières, et en changeant le reste aux zéros.

Troncation des valeurs singulières

- $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots \sigma_n u_n v_n^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$
- On peut approximer **A** en prenant seulement les premières r-colonnes de **U** (et **V**) et les premières r-valeurs singulières, et en changeant le reste aux zéros.
- Avec $r = 2$: $\mathbf{A}_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \sum_{i=1}^{2} \sigma_i u_i v_i^T$

Troncation des valeurs singulières

- $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \ldots \sigma_n u_n v_n^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$
- On peut approximer **A** en prenant seulement les premières r-colonnes de **U** (et **V**) et les premières r-valeurs singulières, et en changeant le reste aux zéros.
- Avec $r = 2$: $\mathbf{A}_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \sum_{i=1}^{2} \sigma_i u_i v_i^T$
- Avec $r = 4$: $\mathbf{A}_4 = \sum_{i=1}^{4} \sigma_i u_i v_i^T$

Troncation des valeurs singulières

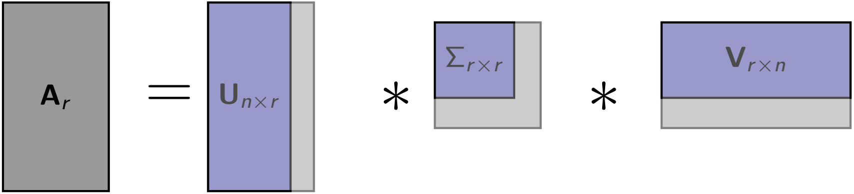
- $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots \sigma_n u_n v_n^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$
- On peut approximer $\mathbf{A}$ en prenant seulement les premières r-colonnes de $\mathbf{U}$ (et $\mathbf{V}$) et les premières r-valeurs singulières, et en changeant le reste aux zéros.
- Avec $r = 2$: $\mathbf{A}_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \sum_{i=1}^{2} \sigma_i u_i v_i^T$
- Avec $r = 4$: $\mathbf{A}_4 = \sum_{i=1}^{4} \sigma_i u_i v_i^T$
- L'erreur d'approximation est donnée par :

$$\text{Erreur}_r = \|\mathbf{A_r} - \mathbf{A}\|_F = \sqrt{\sum_{i=1}^{\min(n,r)} \sigma_i^2} \tag{3}$$

ou $\|X\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |X_{ij}|^2}$ est la norme de Frobenius.

Troncation des valeurs singulières

Le paramètre de troncation r décide le rang de la matrice d'approximation :



$\mathbf{A}_r$ est la meilleure approximation de $\mathbf{A}$ avec le rang r

Utilité de la troncation :

- Compression des données (diminution de l'espace disque)
- De-bruitage (y compris inférence de données manquantes)

Exemple en R : Instructions

- Pour travailler avec des images **install.packages(png)**
- Activer le bibliothèque avec **library(png)**
- Télécharger l'exemple **SVD.R** sur le Moodle
- D'abord lire le code pour comprendre
- Ensuite exécuter le code pour voir les résultats de compression
- Terminer les 3 tâches proposées (**noté**)
- Liens utiles :
  `http://www.statmethods.net/advstats/matrix.html`

Applications

- Réduction de dimensionnalité (ACP) : Factorisation de $X^T X$, Traitement de données et d'images, ...
- Compression et de-bruitage par calcul de rang des données.
- Optimisation des requêtes Google : L'un des plus grands jeux de données numériques est l'Internet. Compression de matrice documents-termes M$(10^6 \times 10^9)$
- Calcul de similarités pour les systèmes de recommandation (séance)
- Évaluation d'inverse d'une matrice arbitraire non carrée (pseudo-inverse)
- Analyse des transformations linéaire et systèmes linéaires

Résumé

- Le SVD est une manière de factoriser une matrice non-carré qui généralise l'opération de décomposition en valeurs propres (pour une matrice symétrique)
- La troncation des valeurs singulières produit des approximation de bas-rang d'une matrice d'entrée.
- Le rang d'un matrice est les nombres des valeurs singulières non-nul.
- Les matrices **U** et **V** sont orthogonales et unitaire : il préserve les distances entres les vecteurs de A.
- La décomposition SVD pour une matrice non-carrée est unique

De-bruitage Exemple

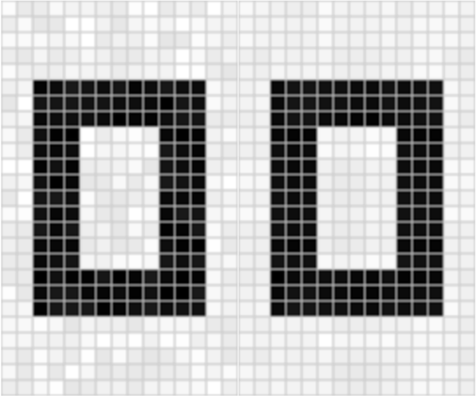


FIGURE – Gauche : Image originale Droit : Image de-bruité (avec 3 valeurs singulières)

Source : [lien]

Valeurs singulières

$\sigma_1 = 14.15$

$\sigma_2 = 4.67$

$\sigma_3 = 3.00$

$\sigma_4 = 0.21$

$\sigma_5 = 0.19$

$\vdots$

$\sigma_{15} = 0.05$