TD : La Décomposition en Valeurs Singulières (SVD)

Rappel Fondamental

Toute matrice $\mathbf{A}$ de taille $m \times n$ peut s'écrire :

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

où σ_i sont les valeurs singulières, ordonnées $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

---

1 Exercice 1 : Rang, Valeurs Singulières et Normes

Soit une matrice $\mathbf{A}$ dont les valeurs singulières non nulles sont données par le vecteur :

$$\mathbf{s} = \begin{pmatrix} 100 \\ 50 \\ 10 \\ 5 \\ 0.1 \end{pmatrix}$$

1. Quel est le **rang** (r) de la matrice $\mathbf{A}$?
2. Rappeler comment est calculée la **norme de Frobenius** $||\mathbf{A}||_F^2$ en utilisant les valeurs singulières σ_i.
3. Calculer la valeur numérique de $||\mathbf{A}||_F^2$.
4. Rappeler comment est calculée la **norme spectrale** $||\mathbf{A}||_2$. Donner sa valeur numérique.

---

2 Exercice 2 : Théorème d'Eckart-Young et Erreur de Troncature

Le Théorème d'Eckart-Young stipule que la meilleure approximation de rang k est la matrice tronquée $\mathbf{A}_k$.

1. Écrire l'expression de l'approximation de rang $k = 3$ de la matrice $\mathbf{A}$ de l'Exercice 1, notée $\mathbf{A}_3$, en utilisant la notation de sommation SVD.
2. D'après la dérivation du Théorème d'Eckart-Young, l'erreur au carré de la reconstruction est donnée par :
$$||\mathbf{A} - \mathbf{A}_k||_F^2 = \sum_{i=k+1}^{r} \sigma_i^2$$
3. Calculer l'erreur de reconstruction $||\mathbf{A} - \mathbf{A}_3||_F^2$ pour la matrice $\mathbf{A}$ de l'Exercice 1 (où $k = 3$).
4. Quel pourcentage de l'information (au sens de la norme de Frobenius au carré) a été perdu lors de cette troncature ?

---

3 Exercice 3 : Lien entre SVD et Analyse en Composantes Principales (ACP)

Soit $\mathbf{X}$ une matrice de données standardisées (n observations $\times$ p variables) et sa SVD $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

1. En ACP, les **composantes principales** (CP) sont les nouvelles variables obtenues après projection. Expliquer comment elles sont dérivées à partir de la SVD. Quelle est la dimension de la matrice des composantes principales (scores) $\mathbf{X}_{proj}$ si l'on ne conserve que k composantes ?
2. Que représentent physiquement les colonnes de la matrice $\mathbf{V}$ (ou $\mathbf{V}_k$) dans le contexte de l'ACP ? (Rappel : ce sont les vecteurs singuliers droits).
3. Si $\sigma_1^2 / \sum \sigma_i^2 = 0.65$, quelle est la signification de cette valeur dans le cadre de l'ACP ?

---

4 Exercice 4 : Application - Dénouement d'Image (Denoising)

Considérons une matrice $\mathbf{A}_{bruite}$ représentant une image. La décomposition SVD donne le vecteur des valeurs singulières suivantes :

$$\mathbf{s}_{bruit} = \begin{pmatrix} 800 \\ 50 \\ 2 \\ 1.5 \\ 1.1 \\ 0.9 \\ \dots \end{pmatrix}$$

1. Expliquer le principe du dénouement d'image (Image Denoising) par SVD. Pourquoi le "signal" (l'image claire) se retrouve-t-il dans les premières valeurs singulières, et le "bruit" dans les suivantes ?

2. Proposer un rang de troncature k approprié pour ce dénouement et justifier votre choix en vous basant sur la magnitude des σ_i.
3. Écrire la formule de la matrice d'image dénouée $\mathbf{A}_{dnoue}$ pour le rang k que vous avez choisi.

---

5 Exercice 5 : Application - PageRank et Vecteurs Propres Dominants

Le PageRank de Google est initialement basé sur le calcul du vecteur propre dominant de la matrice de transition des liens $\mathbf{M}$ (associé à la valeur propre $\lambda = 1$).

1. Expliquer la relation conceptuelle entre le vecteur propre dominant (utilisé pour le PageRank) et les premiers vecteurs singuliers $\mathbf{u}_1$ et $\mathbf{v}_1$ de la SVD.
2. Pourquoi, en pratique sur de très grandes matrices non-symétriques (comme les graphes de liens web), est-il souvent préférable d'utiliser des méthodes itératives basées sur l'itération des puissances (Power Iteration), qui est très proche des calculs SVD/propres, plutôt que de chercher une solution analytique à la Décomposition Spectrale ?