FONDEMENTS DU MACHINE LEARNING

10 septembre 2024

Au programme

- Cours 2 : SVD
- TD 1 : Mise en pratique

Décomposition en valeurs singulières / Singular value decomposition (SVD)

Motivation : Jeu de données $X \in \mathbb{R}^{m \times n}$

But : Comprendre la structure de X pour potentiellement représenter X avec moins de $m \times n$ coefficients

Approche : Analyser X en tant qu'application linéaire

(1) Sous-espaces fondamentaux

Définition : Pour toute matrice $X \in \mathbb{R}^{m \times n}$, on définit :

- le noyau de X : $\ker(X) := \{ v \in \mathbb{R}^n \mid Xv = \overset{0_{\mathbb{R}^m}}{0} \}$
- l'image de X : $\operatorname{Im}(X) := \{ u \in \mathbb{R}^m \mid \exists v \in \mathbb{R}^n, u = Xv \}$

↳ 4 sous-espaces fondamentaux liés à X : $\ker(X), \operatorname{Im}(X)$
$\ker(X^T), \operatorname{Im}(X^T)$

Définition : . Le rang de X est la dimension de $\operatorname{Im}(X)$

Théorème du rang : $\forall X \in \mathbb{R}^{m \times n}$, $\dim(\ker(X)) + \operatorname{rang}(X) = n$

② Cas des matrices symétriques

Dans cette partie, on suppose que $m = n$ et $X^T = X$

Alors, il existe une décomposition spectrale de X de la forme

$$X = P \Lambda P^T$$

$P \in \mathbb{R}^{m \times m}$ orthogonale $(P^T = P^{-1})$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix}$$ matrice diagonale

$\lambda_1 \geq \dots \geq \lambda_m$ réels

valeurs propres

- P permet de transformer X en une matrice diagonale
- Les valeurs propres nous disent comment X transforme les colonnes de P

$\Rightarrow$ Le rang de X est le nombre de valeurs propres non nulles

$\Rightarrow$ Le noyau de X est engendré par les colonnes de P correspondant aux valeurs propres nulles

$\Rightarrow$ Pour $\lambda_i \neq 0$, si $|\lambda_i| \gg 1$, alors X a un effet d'augmentation dans la direction de la i-ème colonne de P

si $|\lambda_i| \ll 1$, alors X a un effet de réduction dans la direction de la $i$ème colonne de P

$P = [p_1 \dots p_m]$

$p_i \in \mathbb{R}^n \ \forall i$

$$X p_i = \lambda_i p_i$$

Ex)

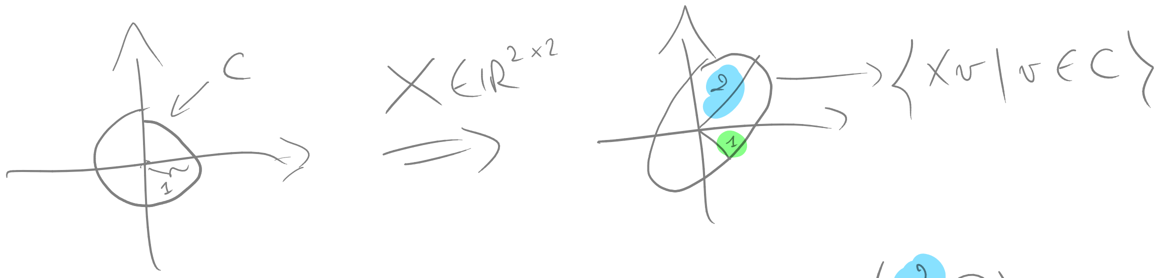
$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}_{P} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}_{P^T = P^{-1}}$$

- 1 valeur propre nulle $\Rightarrow$ rang$(X) = 1$

  On dit que la matrice n'est pas de rang plein, ce qui signifie qu'elle contient moins d'information que sa capacité

- 1 valeur propre égale à 1 $\Rightarrow$ conservation des vecteurs colinéaires à $p_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

Ex)



Avec la bonne base P, X s'écrit $P \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} P^T$

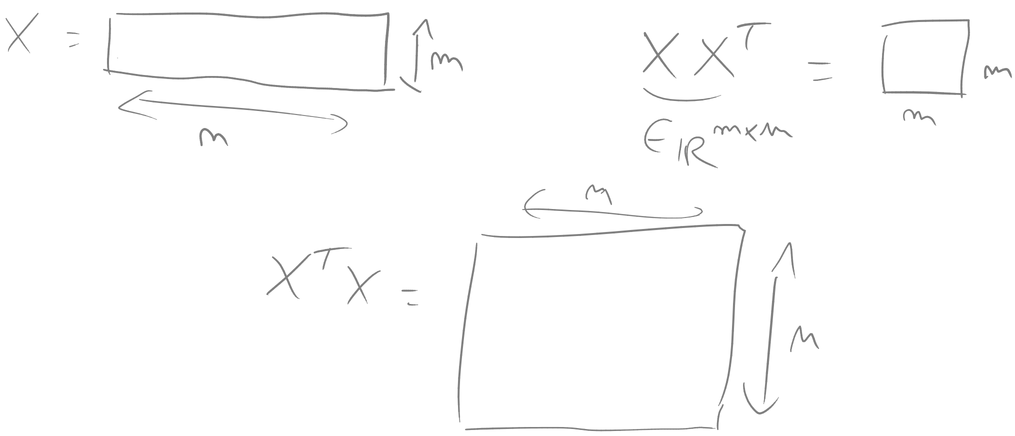
③ Cas général

$\hookrightarrow$ On revient au cas $X \in \mathbb{R}^{m\times n}$ avec m, n quelconques (≥ 1)

$\hookrightarrow$ Q : Existe-t-il un analogue de la décomposition spectrale dans ce cas général ?

Idée : X n'est pas nécessairement symétrique (ni carrée), mais $\underbrace{X^T}_{n\times m}\underbrace{X}_{m\times n}$ et $\underbrace{X}_{m\times n}\underbrace{X^T}_{n\times m}$ le sont !

NB :



Propriétés des matrices XX^T et X^TX

- $\ker(X^TX) = \ker(X)$ et $\ker(XX^T) = \ker(X^T)$
- $\mathrm{Im}(X^TX) = \mathrm{Im}(X^T)$ et $\mathrm{Im}(XX^T) = \mathrm{Im}(X)$
- $\mathrm{rang}(X^TX) = \mathrm{rang}(X)$ et $\mathrm{rang}(XX^T) = \mathrm{rang}(X^T)$

$\Rightarrow$ L'étude de X via ses 4 sous-espaces fondamentaux se ramène à l'étude des matrices symétriques XX^T et X^TX, qui ont chacune 2 sous-espaces fondamentaux.

Cela revient à dire que l'on peut étudier X via les décompositions spectrales de XX^T et X^TX.

- Les valeurs propres de XX^T et X^TX sont positives ou nulles ($XX^T \succeq 0$ et $X^TX \succeq 0$)

Théorème (SVD)

Soit $X \in \mathbb{R}^{m \times n}$. Cette matrice admet une décomposition en valeurs singulières de la forme :

$$X_{m\times n} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T} \quad \text{avec} \quad \begin{array}{l} U \in \mathbb{R}^{m\times m},\ U^T = U^{-1} \ \text{(orthogonale)} \\ V \in \mathbb{R}^{n\times n},\ V^T = V^{-1} \end{array}$$

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_{min(m,n)} & \\ \hline & 0 & & 0 \end{array}\right] \in \mathbb{R}^{m\times n}$$

$\Sigma = (\sigma_{ij})$ avec $\sigma_{ij} = 0$ si $i \neq j$
$\sigma_{ii} = \sigma_i \geq 0$

- Colonnes de U : vecteurs singuliers à gauche
- Colonnes de V : vecteurs singuliers à droite
- Coefficients "diagonaux" de Σ

$\sigma_1, \ldots, \sigma_{min(m,n)}$

valeurs singulières (en général on les ordonne : $\sigma_1 \geq \ldots \geq \sigma_{min(m,n)} \geq 0$)

Ex) $X \in \mathbb{R}^{3\times 2}$

$$X = \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{U} \ \underbrace{\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \ \underbrace{\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}}_{V^T}$$

$u_1 \in \mathbb{R}^3$, $u_2 \in \mathbb{R}^3$, $u_3 \in \mathbb{R}^3$

$\sigma_1 \geq \sigma_2 \geq 0$

$v_1 \in \mathbb{R}^2$
$v_2 \in \mathbb{R}^2$

$X \in \mathbb{R}^{1\times 2}$

$X = [x \ y]$

$$X = \underset{U}{[\pm 1]}\ \underset{\Sigma}{[\sigma_1 \ \ 0]}\ \underbrace{\begin{bmatrix} & \\ & \end{bmatrix}}_{2} \Big\} 2$$

Propriétés de la SVD

i) Le rang de X est le nombre de valeurs singulières non nulles $(\text{rang}(X) \leq \min(m,n))$

ii) Les carrés des valeurs singulières de X

$\sigma_1^2 \geq \ldots \geq \sigma_{min(m,n)}^2$

sont valeurs propres de X^TX et XX^T

iii) Si $X = U\Sigma V^T$ est une SVD de X, alors

$U^T \Sigma^T V$ est une SVD de X^T

TD : Preuve constructive de la SVD

Calcul pratique de la SVD : * Efficace grâce à l'algorithme de Golub & Kahan

* Stable numériquement

4) SVD et réduction de dimension

Soit $X \in \mathbb{R}^{m \times n}$ et $X = U \Sigma V^T$ une SVD de X

avec $\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_{\min(m,n)} & \\ \hline & 0 & & 0 \end{array}\right]$ $\quad \sigma_1 \geq \dots \geq \sigma_{\min(m,n)}$

$U = [u_1 \dots u_m]$ avec $u_i \in \mathbb{R}^m \; \forall i$

$V = [v_1 \dots v_n]$ avec $v_j \in \mathbb{R}^n \; \forall j$

SVD réduite de X

$$X = \underbrace{[u_1 \dots u_r]}_{m \times r} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}}_{r \times r} \underbrace{\begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}}_{r \times n}$$

$r = \text{rang}(X)$

SVD tronquée de X / k-SVD

$$X_k = [u_1 \dots u_k] \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_k^T \end{bmatrix}$$

$k \leq \text{rang}(X)$

↳ Si $r \ll \min(m,n)$, alors la SVD réduite permet de représenter X avec moins de $m \times n$ coefficients

↳ Si $k \ll \min(m,n)$, alors la k-SVD permet d'approcher X avec moins de $m \times n$ coefficients