

ML 08 Dérivation du SVD

Après ces préambules sur la notion de métrique associée à une matrice définie non négative et sur la méthode d'optimisation dite des multiplicateurs de Lagrange, nous reprenons la dérivation de la preuve de l'existence de la décomposition en valeurs singulières pour toute matrice.

Rappelons que nous avons noté \mathbf{v}_1 le vecteur directeur du meilleur sous-espace de dimension 1 pour le nuage des N observations de \mathcal{R}^P . Nous prouvons maintenant que \mathbf{v}_1 est le vecteur propre de $\mathbf{X}^T \mathbf{X}$ associé à la plus grande valeur propre λ_1 .

Le résultat est sans difficulté plus général. Nous le prouvons pour toute matrice symétrique \mathbf{A} (non seulement $\mathbf{X}^T \mathbf{X}$) et toute métrique de \mathcal{R}^P représentée par une matrice symétrique définie non négative \mathbf{M} (non seulement la matrice identité \mathbf{I}).

Nous rappelons que $\mathbf{v}^T \mathbf{A} \mathbf{v} = \sum a_{i,j} v_i v_j$. Ainsi, comme \mathbf{A} et \mathbf{M} sont symétriques (i.e., $a_{i,j} = a_{j,i}$ et $m_{i,j} = m_{j,i}$), les dérivées partielles des formes quadratiques ont les formes suivantes :

$$\frac{\partial \mathbf{v}^T \mathbf{A} \mathbf{v}}{\partial \mathbf{v}} = 2\mathbf{A} \mathbf{v} \quad \text{et} \quad \frac{\partial \mathbf{v}^T \mathbf{M} \mathbf{v}}{\partial \mathbf{v}} = 2\mathbf{M} \mathbf{v}$$

Pour trouver le maximum de $\mathbf{v}^T \mathbf{A} \mathbf{v}$ en intégrant la contrainte unitaire sur \mathbf{v} (i.e., $\mathbf{v}^T \mathbf{M} \mathbf{v} = 1$), nous annulons les dérivées du langrangien \mathcal{L} :

$$\mathcal{L} = \mathbf{v}^T \mathbf{A} \mathbf{v} - \lambda(\mathbf{v}^T \mathbf{M} \mathbf{v} - 1)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{v}} &= 0 \\ &= \{\text{voir calcul des dérivées ci-dessus}\} \\ 2\mathbf{A} \mathbf{v} - 2\lambda \mathbf{M} \mathbf{v} &= 0 \\ &= \\ \mathbf{A} \mathbf{v} &= \lambda \mathbf{M} \mathbf{v} \\ &= \{\mathbf{v}^T \mathbf{M} \mathbf{v} = 1\} \\ \lambda &= \mathbf{v}^T \mathbf{A} \mathbf{v} \end{aligned}$$

Nous découvrons que la valeur du multiplicateur de Lagrange λ est le maximum recherché. Par ailleurs, nous avons :

$$\begin{aligned} \mathbf{A} \mathbf{v} &= \lambda \mathbf{M} \mathbf{v} \\ &= \{\mathbf{M} \text{ est définie non négative et donc inversible}\} \\ \mathbf{M}^{-1} \mathbf{A} \mathbf{v} &= \lambda \mathbf{v} \end{aligned}$$

Donc \mathbf{v} est le vecteur propre de $\mathbf{M}^{-1} \mathbf{A}$ associé à la plus grande valeur propre λ . Nous notons cette solution \mathbf{v}_1 et la valeur propre correspondante λ_1 .

Nous cherchons ensuite \mathbf{v}_2 , unitaire ($\mathbf{v}_2^T \mathbf{M} \mathbf{v}_2 = 1$), orthogonal à \mathbf{v}_1 ($\mathbf{v}_1^T \mathbf{M} \mathbf{v}_2 = 0$) et qui maximise $\mathbf{v}_2^T \mathbf{A} \mathbf{v}_2$. Pour ce faire, nous annulons les dérivées du Lagrangien ci-après.

$$\mathcal{L} = \mathbf{v}_2^T \mathbf{A} \mathbf{v}_2 - \lambda_2 (\mathbf{v}_2^T \mathbf{M} \mathbf{v}_2 - 1) - \mu_2 \mathbf{v}_2^T \mathbf{M} \mathbf{v}_1$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{v}_2} &= 0 \\ &= \{\text{Par définition de } \mathcal{L} \text{ et car } \mathbf{A} \text{ et } \mathbf{M} \text{ sont symétriques.}\} \\ &\quad 2\mathbf{A} \mathbf{v}_2 - 2\lambda_2 \mathbf{M} \mathbf{v}_2 - \mu_2 \mathbf{M} \mathbf{v}_1 = 0 \\ \Rightarrow \{ &\text{Multiplier à gauche chaque membre par } \mathbf{v}_1^T \\ &\quad \mathbf{v}_1^T \mathbf{A} \mathbf{v}_2 = 0 ; \mathbf{v}_1^T \mathbf{M} \mathbf{v}_2 = 0 ; \mathbf{v}_1^T \mathbf{M} \mathbf{v}_1 = 1 \} \\ &\quad \mu_2 = 0 \end{aligned}$$

Nous avons utilisé ci-dessus la propriété $\mathbf{v}_1^T \mathbf{A} \mathbf{v}_2 = 0$ que nous prouvons ci-dessous.

$$\begin{aligned} &\mathbf{v}_1^T \mathbf{A} \mathbf{v}_2 \\ &= \{\mathbf{A} \text{ est symétrique.}\} \\ &\quad \mathbf{v}_2^T \mathbf{A} \mathbf{v}_1 \\ &= \{\mathbf{A} \mathbf{v}_1 = \lambda_1 \mathbf{M} \mathbf{v}_1 \text{ car } \mathbf{v}_1 \text{ est vecteur propre de } \mathbf{M}^{-1} \mathbf{A}\} \\ &\quad \lambda_1 \mathbf{v}_2^T \mathbf{M} \mathbf{v}_1 \\ &= \{\mathbf{v}_1 \text{ et } \mathbf{v}_2 \text{ sont M-orthogonaux.}\} \\ &\quad 0 \end{aligned}$$

Ainsi, nous trouvons :

$$\begin{aligned} &\mathbf{A} \mathbf{v}_2 = \lambda_2 \mathbf{M} \mathbf{v}_2 \\ &= \{\mathbf{M} \text{ est définie non négative donc inversible.}\} \\ &\quad \mathbf{M}^{-1} \mathbf{A} \mathbf{v}_2 = \lambda_2 \mathbf{v}_2 \end{aligned}$$

Donc, \mathbf{v}_2 est le vecteur propre de $\mathbf{M}^{-1} \mathbf{A}$ qui correspond à la seconde plus grande valeur propre λ_2 . Le raisonnement est similaire pour \mathbf{v}_3 , etc.