Recommandation - Rapport partiel

Ken Chanseau–Saint-Germain & Vincent Vidal

16 janvier 2014

Notation

On notera pour $p > 0$ un réel, X un vecteur et M une matrice quelconque :

$$\|X\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}} \qquad \|M\|_p = \left(\sum_{i,j} |m_{i,j}|^p\right)^{\frac{1}{p}}$$

$$|||M|||_p = \sup_{\|x\|_p=1} \|Mx\|_p$$

Et on posera $\|X\|_0$ le nombre de composantes non nulles de X.

1 Introduction

On se donne ici n personnes donnant des notes à m objets.
On notera $a_{i,j}$ la note de l'individu i sur l'objet j ainsi que $A = (a_{i,j})_{\substack{1\leq i\leq n\\1\leq j\leq m}}$ la matrice des notes.

On suppose ici que l'on a accès qu'à une matrice incomplète B obtenu en annulant certaines composantes de A. Le but est alors de trouver une bonne approximation $\widehat{A}$ de A à partir de B.

On prendra comme mesure d'approximation, l'erreur moyenne suivante :

$$\text{RMSE}\left(\widehat{A}\right) = \left\|A - \widehat{A}\right\|_2 = \sqrt{\sum_{i,j} \left(a_{i,j} - \widehat{a}_{i,j}\right)^2}$$

2 Approximation basique

On considère ici l'approximation suivante :

$$\widehat{a}_{i,j} = m_{\widehat{A}} + p_i + o_j$$

Avec $m_{\widehat{A}}$ la moyenne des valeurs de $\widehat{A}$, p_i la moyenne des notes recentrées qu'a donné la personne i et o_j la moyenne des notes recentrées qu'a obtenu l'objet j. C'est à dire :

$$m_{\widehat{A}} = \frac{\sum_{i,j} b_{i,j}}{\|B\|_0} \qquad p_i = \frac{\sum_j b_{i,j} - m_{\widehat{A}}}{\|B^{\mathrm{T}}\mathrm{e}_i\|_0} \qquad o_j = \frac{\sum_i b_{i,j} - m_{\widehat{A}}}{\|B\mathrm{e}_i\|_0}$$