Recommandation - Rapport partiel

Ken Chanseau–Saint-Germain & Vincent Vidal

26 janvier 2014

Table des matières

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 $A = (a_{i,j})_{\substack{1 \le i \le n \\ 1 \le j \le m}}$ la matrice des notes.

On suppose ici que l'on n'a accès qu'à une matrice incomplète B obtenue 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 :

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

En pratique, les données que l'on utilisera pour tester les algorithmes correspondront à des matrices A incomplètes, et on utilisera une matrice B plus incomplète que A. La matrice d'approximation $\widehat{A}$ obtenue donne alors des coefficients dont on ne peut vérifier la précision. L'erreur moyenne ne sera alors calculée que sur les coefficients non nuls de A.

2 Approximation basique

2.1 Par moyenne

On suppose ici que la note d'un objet est de la forme $\widehat{a}_{i,j} = m_{\widehat{A}} + p_i + o_j$, 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 :

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

Algorithm 1: Recommandations naïves par moyenne

Data: La matrice incomplète B.
Result: La matrice d'approximation $\widehat{A}$.

```
m ⟵ mean(B);
for i ∈ ⟦1, n⟧ do
    p[i] ⟵ 0;
    card ⟵ 0;
    for j ∈ ⟦1, m⟧ do
        p[i] ⟵ p[i] + B[i, j];
        if b[i, j] ≠ 0 then
            card ⟵ card + 1;
    p[i] ⟵ p[i]/card − m;
On calcul de même les o[j];
for i ∈ ⟦1, n⟧ do
    for j ∈ ⟦1, m⟧ do
        Â[i, j] ⟵ m + p[i] + o[j];
return Â;
```

2.2 Par minimisation de problème convexe

De manière plus précise, on peut résoudre le problème suivant de minimisation convexe pénalisée sur les p_i et o_j pour trouver des p_i et o_j plus adaptés :

$$\mathcal{P}(p,\, o) = \sum_{i,j} \left(a_{i,j} - m_{\widehat{A}} + p_i + o_j\right)^2 + \lambda \left(\|o\|_1 + \|p\|_1\right)$$

Problème que l'on peut écrire :

$$\mathcal{P}(x) = \|Mx + b\|_2^2 + \lambda \|x\|_1 .$$

Il suffit alors de résoudre ce problème, pour un λ donné et de calculer l'approximation à partir des coefficients p_i et o_j obtenus.

On utilisera ici un algorithme de minimisation qui descend de l'algorithme de Douglas-Rachford [**?**], se rapprochant d'une méthode de descente de gradient. Cet algorithme prend un paramètre supplémentaire, μ, dont dépend la vitesse de convergence. La suite calculée par cet algorithme converge vers un minimum lorsque μ est suffisamment proche de 0.

Algorithm 2: Minimisation convexe

Data: x, λ, le nombre d'itération N et un paramètre μ.
Result: une meilleur approximation x.
for $k \in [\![1,\ N]\!]$ **do**
  $x \longleftarrow x - \mu.M^{\mathrm{T}}\,(Mx + b)$;
  `/* Seuillage doux sur chaque composante */`
  for $i \in [\![1,\ n]\!]$ **do**
    $x\,[i] \longleftarrow \mathrm{signe}\,(x\,[i])\,.\max\,(\ |x\,[i]| - \mu\lambda,\ 0\ )$;
return x;

3 Approximation par voisinage

Nous noterons par la suite P_i la i-ième colonne de B^{T}, représentant la personne i et O_j la j-ième colonne de B, représentant l'objet j.

3.1 Plus proche voisin

La méthode précédente suppose qu'il n'existe qu'un seul type de personne et d'objet : un objet est bien ou mauvais dans l'ensemble et une personne donne de manière générale des notes bonnes ou des notes mauvaises. Ceci apparaît comme très naïf.

On considère ici une méthode différente. Pour une personne i et un objet j dont l'information est manquante, nous allons chercher parmi les personnes ayant noté j celui qui se rapproche le plus i ainsi que l'objet ayant été noté par i celui qui se rapproche le plus de j. En notant i_0 et j_0 respectivement la personne et l'objet trouvés, on peut alors prendre comme approximation la moyenne des deux notes : $\widehat{a}_{i,j} = \frac{1}{2}\,(a_{i_0,j} + a_{i,j_0})$.

Il reste à définir comment sélectionner le plus proche voisin. Si on cherche le plus proche voisin d'un vecteur X de support S parmi un ensemble de vecteurs E, on considérera $E_{|S}$ les vecteurs de E projetés sur le support S et on prendra alors le plus proche voisin pour la norme euclidienne sur $\mathbb{R}^{\mathrm{Card}(S)}$.

3.2 k-plus proches voisins

On peut vouloir effectuer l'algorithme précédent en moyennant les notes des k-plus proches voisins. L'algorithme est en tout point semblable au précédent.

3.3 Voisinage

On ne cherche plus ici le plus proche voisin, mais une combinaison de la note donnée par les personne ayant noté l'objet en question, et ce en fonction de leur ressemblance avec la personne initiale.

Il convient de donner une mesure de ressemblance entre deux personnes i_1 et i_2, que l'on notera $s(i_1,\ i_2)$. Nous prendrons ici la valeur du cosinus de l'angle entre les deux personnes :

$$s(i_1,\ i_2) = \frac{P_{i_1}^{\mathrm{T}} \cdot P_{i_2}}{\|P_{i_1}\|_2 \, \|P_{i_2}\|_2}.$$

On prendra alors comme approximation :

$$\widehat{a}_{i_0,j} = \frac{\sum\limits_i \left(s(i_0,\ i) \cdot b_{i,j}\right)}{\sum\limits_i |s(i_0,\ i)|}.$$

Algorithm 3: Minimisation par voisinage

Data: La matrice B.
Result: La matrice approchée $\widehat{A}$.

$S \longleftarrow BB^{\mathrm{T}}$;
for $i \in [\![1,\ n]\!]$ **do**
  $norm[i] \longleftarrow \left\|B^{\mathrm{T}}\mathrm{e}_i\right\|_2$
for $i \in [\![1,\ n]\!]$ **do**
  $somme[i] \longleftarrow 0$;
  for $k \in [\![1,\ n]\!]$ **do**
    $S[i,\ k] \longleftarrow S[i,\ k] / (norm[i].norm[k])$;
    $somme[i] \longleftarrow somme[i] + |S[i,\ k]|$;
$\widehat{A} \longleftarrow S.B$;
for $i \in [\![1,\ n]\!]$ **do**
  for $j \in [\![1,\ m]\!]$ **do**
    $\widehat{A}[i,\ j] \longleftarrow \widehat{A}[i,\ j] / somme[i]$;
return $\widehat{A}$;

4 Décomposition en valeurs singulières

Le but est ici de faire apparaître k modèles de personnes, dont toutes les autres seront des combinaisons linéaires.

Autrement dit, on souhaite faire apparaître la matrice $\widehat{A}$ de la manière suivante :

$$\widehat{A} = P \cdot O$$

avec $P \in \mathcal{M}_{n,\,k}(\mathbb{R})$ de colonnes libres que l'on peut prendre orthonormées et $O \in \mathcal{M}_{k,\,m}(\mathbb{R})$.

On peut facilement accéder à une telle décomposition à l'aide de la décomposition en valeurs singulières.

Algorithm 4: Recommandation par décomposition en valeurs singulières

---

Data: La matrice B et un entier k.
Result: La matrice approchée $\widehat{A}$.
$[U,\, S,\, V] \longleftarrow \mathrm{SVD}\,(B)$;
for $i \in [\![k,\, \mathrm{size}\,(S)]\!]$ **do**
  $S\,[i,\, i] \longleftarrow 0$;
$\widehat{A} \longleftarrow USV^{\mathrm{T}}$;
return $\widehat{A}$;
