Approximation par les moindres-carrés

1 Introduction

Les problèmes de moindres carrés se rencontrent fréquemment dans des problèmes d'estimation de paramètres (modélisation statistique, ajustement de données, régression linéaire...).

Pour les moindres carrés linéaires, le problème consiste en général à résoudre le système

$$Ax = b \text{ avec } A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m,\ m \geq n.$$

On parle ici de **système linéaire sur-determiné** (car on a plus d'équations que d'inconnues).

Une solution existe si $b \in ImA$, ce qui n'est vrai que dans des cas exceptionnels (car $b \in \mathbb{R}^m$ et $dim(ImA) \leq n$). L'idée des moindres carrés est alors de trouver x tel que $\|Ax-b\|_2$ est le plus petit possible. Le problème s'écrit alors

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2 . \tag{1}$$

2 Exemple

Problème: on a des **observations b_i, $1 \leq i \leq m$ aux points t_i (temps)** et l'on veut estimer (ou prédire) les valeurs de b aux points $t \neq t_i$.

On souhaite alors trouver l'équation $y = f(t)$ qui approche au mieux les points (t_i, b_i) i.e telle que la distance des (t_i, b_i) à la courbe $y = f(t)$ est minimale.

Si par exemple on cherche $f(t)$ sous la forme $\alpha + \beta t$ (droite), alors on détermine α et β tels que $\sum_{i=1}^m \epsilon_i^2$ est minimum avec $\epsilon_i = |f(t_i) - b_i|$. La fonction à minimiser est alors:

$$g(\alpha,\beta) = \sum_{i=1}^m \epsilon_i^2 = \sum_{i=1}^m (\alpha + \beta t_i - b_i)^2.$$

Pour minimiser g, on annule ses dérivées partielles:

$$\left\{ \begin{array}{rcl} \frac{\partial g}{\partial \alpha}(\alpha,\beta) & = & 2\sum_{i=1}^m (\alpha+\beta t_i - b_i) \quad = \quad 0 \\ \frac{\partial g}{\partial \beta}(\alpha,\beta) & = & 2\sum_{i=1}^m (\alpha+\beta t_i - b_i)t_i \quad = \quad 0 \end{array} \right. .$$

Cela donne le système linéaire suivant, avec pour inconnues α et β:

$$\left\{ \begin{array}{rcrcl} m\,\alpha & + & (\sum_{i=1}^m t_i)\,\beta & = & \sum_{i=1}^m b_i \\ (\sum_{i=1}^m t_i)\,\alpha & + & (\sum_{i=1}^m t_i^2)\,\beta & = & \sum_{i=1}^m t_i b_i \end{array} \right. . \tag{2}$$

Si on pose $A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}$ et $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$, alors $A^T = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{pmatrix}$ et on obtient:

$A^TA = \begin{pmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}$ et $A^Tb = \begin{pmatrix} \sum b_i \\ \sum t_i b_i \end{pmatrix}$.

En notant $x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, le système (2) peut donc s'écrire

$$A^T A\, x \;=\; A^T b, \tag{3}$$

appelé aussi systèmes des **équations normales**.

On peut montrer que, si $A \in \mathbb{R}^{m\times n}$, alors les équations normales ont une solution unique si et ssi $rang(A) = n$ et que dans ce cas la matrice $A^T A$ est symétrique définie positive.

La solution de (3) s'écrit alors $x = \underbrace{(A^T A)^{-1} A^T}_{A^\dagger} b = A^\dagger b$ où l'opérateur $A^\dagger$ est appelé **pseudo-inverse** de A.

Dans notre exemple les t_i sont supposés distincts. Par conséquent $rang(A) = 2$ et on a donc une solution unique $x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$.

De manière générale la solution de (3) s'obtient par la résolution du système de la manière la plus adaptée en fonction de la taille et de la structure de $A^T A$ et de la précision numérique demandée (méthode directe via des factorisations, méthode itérative de type gradient conjugué,...) ou par simples substitutions dans le cas de petits systèmes. Dans le cas présent (matrice 2×2) on peut par exemple utiliser la formule explicite de l'inverse de $A^T A$:

$$x = \underbrace{\frac{1}{det(A^T A)} (com(A^T A))^T}_{(A^T A)^{-1}} A^T b = \frac{1}{m\sum t_i^2 - (\sum t_i)^2} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{pmatrix} \begin{pmatrix} \sum b_i \\ \sum t_i b_i \end{pmatrix}.$$

On observe aussi que:

$\sum_{i=1}^m \epsilon_i^2 = \sum_{i=1}^m (\alpha + \beta t_i - b_i)^2 = (Ax-b)^T(Ax-b) = \|Ax-b\|_2^2$ (à vérifier en exercice).

Le vecteur $r = Ax - b$ est appelé le **résidu** du problème de moindres carrés et $\|r\|_2$ permet d'évaluer l'erreur d'approximation.

Interprétation statistique: En posant $\bar{t} = \frac{\sum t_i}{m}$, $\bar{t^2} = \frac{\sum t_i^2}{m}$, $\bar{b} = \frac{\sum b_i}{m}$ et $\bar{tb} = \frac{\sum t_i b_i}{m}$, le système (2) peut s'écrire (en divisant chaque équation par m):

$$\left\{ \begin{array}{rcrcl} \alpha & + & \bar{t}\,\beta & = & \bar{b} \\ \bar{t}\,\alpha & + & \bar{t^2}\,\beta & = & \bar{tb} \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} \alpha & = & \bar{b} - \beta\bar{t} \\ \underbrace{(\bar{t^2} - \bar{t}^2)}_{var(t)}\beta & = & \underbrace{\bar{tb} - \bar{t}\bar{b}}_{cov(t,b)} \end{array} \right.,$$

d'où

$$\left\{ \begin{array}{rcl} \alpha & = & \bar{b} - \frac{cov(t,b)}{var(t)}\bar{t} \\ \beta & = & \frac{cov(t,b)}{var(t)} \end{array} \right. .$$

$y = \alpha + \beta t$ définit alors l'équation de la **droite de régression linéaire d'un nuage de points**. $(A^T A)^{-1}$ est parfois appelée **matrice de covariance**.

3 Méthodes directes de résolution

On décrit ci-dessous 2 méthodes courantes de résolution du problème de moindres carrés (1).

3.1 Equations normales

Si A est de rang plein (i.e $rang(A) = n$) alors la matrice $A^TA \in \mathbb{R}^{n\times n}$ est symétrique définie positive et admet une factorisation de Cholesky $A^TA = R^TR$ avec R triangulaire supérieure. On résout alors $R^TR\,x = A^Tb$.

Méthode directe de résolution des moindres carrés via les équations normales:

1. Former la matrice A^TA et le vecteur A^Tb (coût $\simeq mn^2$ flops si on exploite la symétrie de A^TA).
2. Calculer la factorisation de Cholesky $A^TA = R^TR$ (coût $\simeq n^3/3$ flops).
3. Résoudre $R^Ty = A^Tb$ puis $Rx = y$ (coût $\simeq 2n^2$ flops).

Le coût calculatoire global sera donc: $mn^2 + n^3/3 + \mathcal{O}(n^2) \simeq mn^2$ flops (si $m \gg n$).

3.2 Factorisation QR

La factorisation QR de A est donnée par

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$$

où Q est une matrice $m\times m$ orthogonale et R est une matrice $n\times n$ triangulaire supérieure. Si A est de rang plein, alors on peut montrer que R est inversible.

Si l'on écrit $Q = \begin{pmatrix} Q_1 , Q_2 \end{pmatrix}$ où Q_1 and Q_2 correspondent respectivement aux n premières colonnes et aux $m-n$ colonnes restantes de Q, alors on a $A = Q_1R$ (appelée aussi factorisation QR "réduite" de A). La quantité à minimiser s'écrit donc:

$\|Ax-b\|_2 = \left\|Q^TAx - Q^Tb\right\|_2$ (car Q^T est orthogonale et donc conserve la norme) avec $Q^TAx = \underbrace{Q^TQ}_{I}\begin{pmatrix} R \\ 0 \end{pmatrix} x = \begin{pmatrix} Rx \\ 0 \end{pmatrix}$ et $Q^Tb = \begin{pmatrix} Q_1^Tb \\ Q_2^Tb \end{pmatrix}$.

D'où $\|Ax-b\|_2^2 = \left\| \begin{matrix} Rx - {Q_1}^Tb \\ {Q_2}^Tb \end{matrix} \right\|_2^2 = \left\|Rx - Q_1^Tb\right\|_2^2 + \left\|Q_2^Tb\right\|_2^2$, qui sera minimum lorsque $\left\|Rx - Q_1^Tb\right\|_2 = 0$ i.e $Rx = {Q_1}^Tb$ qui est un système triangulaire supérieur qu'il conviendra de résoudre.

Méthode directe de résolution des moindres carrés via la factorisation QR:

1. Calculer la factorisation réduite $A = Q_1R$.
2. Calculer le vecteur Q_1^Tb.
3. Résoudre $Rx = {Q_1}^Tb$.

Le coût calculatoire sera environ: $2mn^2 - 2n^3/3 \simeq 2mn^2$ flops (si $m \gg n$).

Remarque: en dépit de son coût calculatoire 2 fois plus élevé, la méthode via QR (basée sur des transformations de Householder) est souvent privilégiée pour des raisons de précision.

3.3 Exercice

Une petite entreprise exerce son activité depuis 4 ans et son chiffre d'affaires (en dizaine de milliers d'euros) s'établit comme suit:

année	1	2	3	4
ventes	23	27	30	34

1. Peut-on raisonnablement envisager une tendance linéaire de l'augmentation des ventes?

2. Prédire les ventes pour les années futures si la tendance se confirme (on utilisera la méthode des équations normales).

3. Calculer l'erreur commise par cette approximation.