Méthodes aux moindres carrés, des données aux modèles

1ère partie : moindres carrés linéaires

Exposé par Jérôme MONNIER, Professeur à l'INSA de Toulouse, Département de Génie Mathématique et Modélisation.

Plan :

- Des données à représenter / modéliser,
- Solution au sens des moindres carrés,
- Equations normales,
- Exemples,
- Interprétation géométrique.

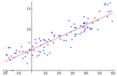
Par la suite, **partie II** :

- Les moindres carrés non linéaires et le choix des fonctions de base,
- La résolution numérique : l'algorithme de Gauss-Newton et la méthode de Householder.

Des données à représenter / modéliser

Soient m données $d_{1\leq i\leq m}$. But : les représenter par un *modèle*.

Ex. *modèle dit de régression linéaire* : une droite : 2 paramètres



Modèle *linéaire à* n paramètres $(x_i)_{1\leq i\leq n}$:

$$\boxed{\left\{\begin{array}{rcl} a_{11}x_1+\ldots+a_{1n}x_n & = & d_1 \\ \ldots & = & \ldots \\ a_{m1}x_1+\ldots+a_{mn}x_n & = & d_m \end{array}\right.}$$

$\rightarrow$ A matrice *rectangulaire* $m\times n$, $d\in\mathbb{R}^m$ vecteur des obs. et $x\in\mathbb{R}^n$ vecteur des inconnues.

$\boxed{\textit{Etant donné } d\in\mathbb{R}^m\textit{, trouver } x\in\mathbb{R}^n \textit{ tel que : } Ax=d.}$

$\rightarrow$ Problème de *calibration de modèle*.

Des données à représenter / modéliser

Problème : résoudre

$$Ax = d$$

avec A matrice *rectangulaire* $m \times n$, $d \in \mathbb{R}^m$ donné et $x \in \mathbb{R}^n$ vecteur inconnu.

- Cas (improbable) $n = m$: pb *bien posé* ssi A est de rang maximal i.e. $r = rank(A) = dim(Im(A)) = n$.
$\rightarrow \exists$ *unique* jeu de paramètres x décrivant *exactement* les données.

- Cas $n = m$ mais $r = rank(A) = dim(Im(A)) < n$,
alors $Ker(A) \neq \emptyset$: $\exists z \in \mathbb{R}^n$ tq $Az = 0$ dans $\mathbb{R}^m$.

- Cas $n > m$: système *sous-déterminé*, infinité de solutions.

- Cas $n < m$, système *sur-déterminé* : cas qui nous intéresse.
A-priori le système n'admet *pas* de solution...

Exercice. Illustrer succintement les cas évoqués, en faisant des figures dans $\mathbb{R}^3$.
1) Etant donné un ensemble de $m = 10$ points de $\mathbb{R}^3$, et un modèle linéaire à $n = 3$ paramètres : cas sur-déterminé.
2) Etant donné un ensemble de $m = 2$ points de $\mathbb{R}^3$, et un modèle linéaire à $n = 3$ paramètres : cas sous-déterminé.

Des données à représenter / modéliser : un exemple

Positions d'un des missiles passé et relevées par un radar :

x_i	0	250	500	750	1000
y_i	0	8	15	19	20

L'a-priori considéré : trajectoire parabolique
$\Rightarrow y(x) = a_0 + a_1\, x + a_2\, x^2 \rightarrow$ 3 inconnues : $a_0,\ a_1$ et a_2.

Système linéaire sur-déterminé de $m = 5$ équations et $n = 3$ inconnues :

$$\left\{\begin{array}{ccccccc} a_0 & & & & & = & 0, \\ a_0 & + & 2.5 a_1 & + & 6.25\, a_2 & = & 0.08, \\ a_0 & + & 5\, a_1 & + & 25\, a_2 & = & 0.15, \\ a_0 & + & 7.5\, a_1 & + & 56.25\, a_2 & = & 0.19, \\ a_0 & + & 10\, a_1 & + & 100\, a_2 & = & 0.20. \end{array}\right.$$

Solution au sens des moindres carrés

On chercher une solution qui représente *au mieux* les observations i.e. *minimisant la norme* $\|Ax - d\|$.
Bon choix de norme : $\|.\|_2$ car *associée à un produit scalaire*.
Le problème devient alors :

Trouver $x^* \in \mathbb{R}^n$ tel que :

$$j(x^*) = \min_{x \in \mathbb{R}^n} j(x) \text{ avec } j(x) = \|Ax - d\|_{2,m}^2$$

$\rightarrow$ *Problème d'optimisation* (dans l'espace entier $\mathbb{R}^n$).
$\rightarrow$ *Solution au sens des Moindres Carrés.*

$j : \mathbb{R}^n \rightarrow \mathbb{R}\,;\, j(x) =< A^T Ax, x > -2 < Ax, d > + < d, d >.$

La matrice carrée $A^T A$ est positive, j est quadratique donc convexe $\Rightarrow$ j admet un minimum (au moins) dans $\mathbb{R}^n$.

Si de plus la matrice $A^T A$ est définie alors la solution est unique.

Solution au sens des Moindres Carrés : équations normales

Rappel C.N. d'optimum local : $\nabla j(x) = 0$.
puis C.S. de minimum (local) : $H_j(x)$ est positive définie.

Lemme. Son gradient : $\nabla j(x) = A^T A x - A^T d$ (vecteur de $\mathbb{R}^n$).
Sa matrice Hessienne (carrée $n \times n$) : $H_j(x) = A^T A$.

Preuve. D.L. de Taylor ordre 2 au voisinage du point x :

$$\begin{aligned}
j(x+h) &= \langle A\,(x+h) - d,\; A\,(x+h) - d\rangle\,, \\
&= \langle (A\,x - d) + A\,h,\; (A\,x - d) + A\,h\rangle\,, \\
&= j(x) + \langle A\,h,\; A\,x - d\rangle + \langle A\,x - d,\; A\,h\rangle + \langle A\,h,\; A\,h\rangle, \\
&= j(x) + 2\,\langle A^T\,(A\,x - d),\; h\rangle + \langle A^T\,Ah,\; h\rangle.
\end{aligned}$$

On reconnait la forme :
$f(x+h) = f(x) + \langle \nabla f(x),\; h\rangle + \frac{1}{2}\,\langle H_f(x)\,h,\; h\rangle + \|h\|_2^2\,\varepsilon(h).$
On en déduit alors les expressions de $\nabla j(x)$ et $H_j(x)$.

Solution au sens des M.C. : équations normales (suite)

Théorème. Toute solution du problème de M.C., cas sur-déterminé $n < m$, est également solution du système linéaire $n \times n$:

$$\boxed{A^T A x = A^T d}$$

$\rightarrow$ Equations normales.
La solution est unique ssi A de rang maximal i.e. $r = rang(A) = n$.

Preuve. C.N. d'optimum local $\nabla j(x) = 0$ donne directement les équations normales.
La matrice $A^T A$ est symétrique positive, cet extremum est alors un minimum global. Et ce minimum global est unique si $A^T A$ est définie, soit A de rang maximal.

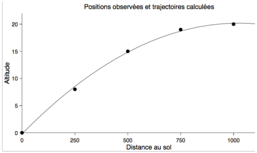
Retour à l'exemple

Les équations normales s'écrivent :

$$\begin{bmatrix} 5 & 25 & 187.5 \\ 25.00 & 187.5 & 1562.5 \\ 187.50 & 1562.5 & 13828.125 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.62 \\ 4.375 \\ 34.9375 \end{bmatrix}$$

$\rightarrow$ Système de 3 équations à 3 inconnues.

Méthode d'élimination de Gauss : solution
$x^* \simeq [-0.0023; 0.0398; -0.0019]$



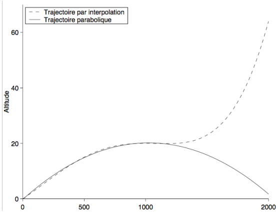
$\rightarrow$ Vérification graphique a-posteriori : la parabole optimale (au sens des M.C.) approche bien les mesures faites par le radar.

Exemple : changeons de modèle pour voir...

Supposons que le défenseur *ne savait a-priori pas* que la trajectoire était parabolique...

5 données $\rightarrow$ unique polynôme de degré 4, $p_4(x)$ (interpolation de Lagrange).

Figure : trajectoire décrite par ce polynôme p_4 sur une distance au sol de 2000 km (i.e. entre 1000 et 2000, il s'agit de valeurs extrapolées).



*M*oindres carrés vs interpolation : le missile s'échappe vers le ciel...

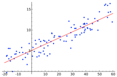
Morale : mauvais modèle calibré $\Rightarrow$ extrapolation / prédiction pas possible...

Méthode des moindres carrés, c'est génial et pas nouveau...

Historiquement : J.C. Gauss (XVIII ième siècle), à 24 ans : prévision précise de la trajectoire d'une asteroïde sur la base de quelques observations (et sur l'hypothèse à l'époque originale d'une trajectoire elliptique) !

Exemple aussi basique qu'utile : modèle de régression lineaire.

Cas le plus simple : faire passer une droite "au mieux" dans un "nuage" de points.



Fonctions de base des solutions : $\{1, x\}$.

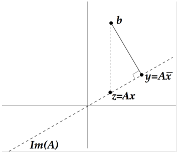
Pour m points (x_i, y_i) de $\mathbb{R}^2$, A s'écrit : $\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}$

Equations normales : $\begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix}$

$\Rightarrow$ Coefficients de la droite (a_0, a_1).

Interprétation géométrique

Solution aux M.C. x^* minimise la distance euclidienne entre $d \in \mathbb{R}^m$, et $Im(A)$, sous-espace vectoriel de $\mathbb{R}^m$.
Soit t.q. Ax^* projeté $\perp$ de $d \in \mathbb{R}^m$ sur $Im(A)$.



" $\|Ax - d\|_{2,m}$ minimal ssi $(Ax - d) \perp Im(A)$"
Démo. Cela est équivalent à : "résidu $r = (Ax - d) \in Im(A)^T$".

On a :
$\forall x \in \mathbb{R}^n, \ \|Ax - d\|_{2,m}^2 = \|Ax^* - d\|_{2,m}^2 + \|Ax - Ax^*\|_{2,m}^2 \geq \|Ax^* - d\|_{2,m}^2$.

$$\begin{aligned} \text{Et } r \in \mathrm{Im}(A)^\perp \quad & \Leftrightarrow \quad \forall y \in \mathrm{Im}(A), \ r^T y = 0, \\ & \Leftrightarrow \quad \forall z \in \mathbb{R}^n, \ r^T A z = 0, \\ & \Leftrightarrow \quad \forall z \in \mathbb{R}^n, \ z^T A^T r = 0, \\ & \Leftrightarrow \quad A^T r = 0 \Leftrightarrow r \in \mathrm{Ker}(A^T) \end{aligned}$$

$\Leftrightarrow$ Equations normales : $A^T(Ax^* - b) = 0$.