Notes de cours
Algèbre Linéaire et Analyse Matricielle
3BIM INSA Lyon
2017-2018
v27112017

Ces notes de cours sont inspirées des livres de Algèbre des matrices de J Fresnel [1] et Numerical Linear Algebra de LN Trefethen et D Bau III [2].

4 Regression linéaire et moindres carrés

Supposons qu'on ait un jeu de données de m observations $y_i \in \mathbb{R}$ pour m conditions $x_i \in \mathbb{R}$, $i = 1, ..., m$. On aimerait approximer le vecteur y comme fonction de x: $y \sim f(x; c)$, où f dépend *linéairement* de n coefficients $c_1, c_2, ...c_n$. Le problème de régression consiste à déterminer les n coefficients qui minimiseront (dans un sens à préciser) la distance entre y et la prédiction $f(x)$.

D'un point de vue algérique, le problème de régression se ré-écrit de la façon suivante. On aimerait trouver une solution à un système sur-déterminé

$$Ac = y \tag{1}$$

où A est une matrice $m \times n$ et $m > n$ qui dépend des données conditions$x \in \mathbb{C}^m$, $c \in \mathbb{C}^n$ est le vecteur de coefficients inconnus et $y \in \mathbb{C}^m$ est le vecteur de données observations. En général, il n'existe pas de vecteur c tel que $Ac = y$ (l'équation telle que posée n'a pas de solution). L'image de A est un s.e.v. de dimension au plus égale à n, et en général y n'en fait pas partie, $y \notin \operatorname{im} A$. Le problème des moindres carrés cherche à trouver le vecteur de coefficients c qui minimise la distance (euclidienne) qui sépare Ac de y, c.-à-d. trouver c qui minimise la norme euclidienne des résidus $r = y - Ac$.

Définition 1. *Problème des moindres carrés.* Étant donné $A \in \mathbb{C}^{m \times n}$, $m > n$, $y \in \mathbb{C}^m$, trouver $c \in \mathbb{C}^n$ tel que $||y - Ac||_2$ est minimisé.

Exemples

Interpolation polynomiale. Soit m points distincts $x_1, x_2, ..., x_m \in \mathbb{C}$, et m données $y_1, y_2, ..., y_m \in \mathbb{C}$. Alors il existe un unique polynôme de degré $m - 1$: $p(x) = c_0 + c_1 x + ... + c_{m-1} x^{m-1}$ tel que $p(x_i) = y_i$ pour tout $i = 1, ..., m$. Le système qui décrit la relation entre x_i et y_i est donné par la matrice de Vandermonde:

$$\begin{pmatrix} 1 & x_1 & x_1^2 & ... & x_1^{m-1} \\ 1 & x_2 & x_2^2 & ... & x_2^{m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & ... & x_m^{m-1} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m-1} \end{pmatrix}. \tag{2}$$

Pour déterminer les coefficients c_i, on peut résoudre ce système d'équations linéaires, qui est non-singulier pour des valeurs x_i distinctes.

Ajustement polynomial par moindres carrés. Si maintenant on prend un polynôme de degré $n-1$ inférieur à $m-1$, $p(x) = c_0 + c_1 x + ... + c_n x^{n-1}$. Pour ajuster ce polynôme au donnés dans le sens des moindres carrés, on veut minimiser la norme euclidienne des résidus $r_i = p(x_i) - y_i$, c.-à-d. minimiser

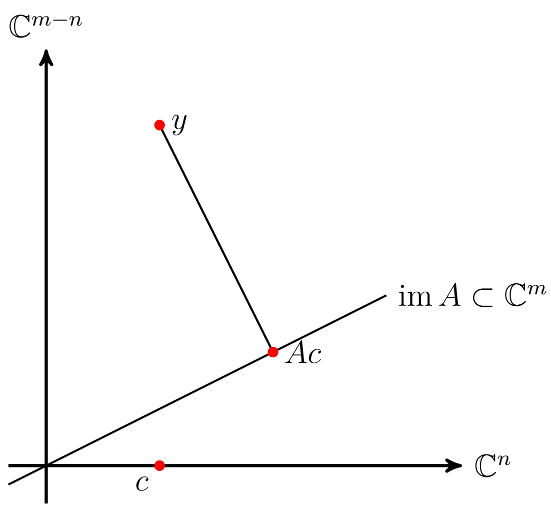
$$\sum_{i=1}^{m} |p(x_i) - y_i|^2. \tag{3}$$

Le système linéaire devient

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \dots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} c_0 \\ \vdots \\ c_{n-1} \end{pmatrix} \sim \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m-1} \end{pmatrix}. \tag{4}$$

Ce système est sur-déterminé (m équations pour n inconnues). Comment trouver c qui minimise r?

Revenons au problème $Ac = y$. L'objectif est de trouver c tel que Ac soit le plus près de y. Géométriquement, on voit que le vecteur de résidus r est minimisé (au sens des moindres carrés) si Ac est la projection orthogonale de y sur $\operatorname{im} A$, c.-à-d. si $Ac = Py$ où P est le projecteur orthogonal sur $\operatorname{im} A$. En d'autres termes, le vecteur résidus r doit être orthogonal à $\operatorname{im} A$!



Théorème 4.1. *Soit une matrice $A \in \mathbb{C}^{m\times n}$, $m > n$, et une donnée $y \in \mathbb{C}^m$. Un vecteur $c \in \mathbb{C}^n$ minimise la norme des résidus $||r||_2 = ||y - Ac||_2$, résolvant ainsi le problèmes des moindres carrés, si et seulement si r est orthogonal à $\operatorname{im} A$, c.-à.d. si*

$$A^* r = 0 \tag{5}$$

ou bien

$$A^*(Ac) = A^*y \quad (\textit{équations normales}) \tag{6}$$

On note que r est orthogonal à $\operatorname{im} A$ si et seulement si pour tout $z \in \mathbb{C}^n$, $(Az)^*r = 0$, ou encore $z^*(A^*r) = 0$, ce qui implique $A^*r = 0$. Les équations normales s'obtiennent en prenant $r = y - Ac$.

La matrice A^*A est une matrice $n \times n$ et est inversible si et seulement si $\operatorname{rang} A = n$. Le problème des moindres carrés à donc une solution unique si et seulement si $\operatorname{rang} A = n$. en ce cas, la solution est donnée par les équations normales:

$$c = (A^*A)^{-1}A^*y. \tag{7}$$

La matrice $(A^*A)^{-1}A^*$ est appelée **pseudo-inverse** de A, et est dénotée A^+.

A^+ est une matrice $n \times m$. Elle envoie les vecteurs $y \in \mathbb{C}^m$ vers les vecteurs $c \in \mathbb{C}^n$. Le problème des moindres carrés revient à calculer

$$c = A^+y \tag{8}$$

où A^+ est la pseudo-inverse de A.

4.1 Algorithme des moindres carrés par décomposition en valeurs singulières

Soit A une matrice $m \times n$ avec $m > n$. On suppose que $\operatorname{rang} A = n$. On construit une décomposition en valeurs singulières, $A = \hat{U}\hat{\Sigma}V^*$, avec $\hat{U}$ une matrice $m \times n$, $\hat{\Sigma}$ une matrice diagonale de valeurs singulières strictement positives et V une matrice unitaire $n \times n$.

La matrice $\hat{U}$ est composée des vecteurs orthonormaux des semi-axes $\sigma_i u_i = Av_i$ de l'ellipse image de la sphère sous A. De plus $\hat{U}\hat{U}^*$ est un projecteur orthogonal, qui projète $y \in \mathbb{C}^m$ sur $\operatorname{im} A$. Les n colonnes de $\hat{U}$, dénotées u_i forment une base orthonormée de $\operatorname{im} A$. Pour $z \in \operatorname{im} A$,

$$\begin{aligned} z &= \sum_{i=1}^{n} (u_i^* z) u_i && (9)\\ &= \sum_{i=1}^{n} u_i (u_i^* z) && (10)\\ &= \sum_{i=1}^{n} (u_i u_i^*) z && (11)\\ &= \hat{U}\hat{U}^* z. && (12) \end{aligned}$$

Le vecteur y se décompose en une partie dans l'image de A et un partie orthogonale r: $y = Ac + r$. En applicant le projecteur de chaque côté, on obtient $\hat{U}\hat{U}^* y = Ac$.

De la décomposition en valeurs singulières

$$A^*Ac = A^*y \tag{13}$$

se réécrit

$$(\hat{U}\hat{\Sigma}V^*)^*\hat{U}\hat{\Sigma}V^*x = (\hat{U}\hat{\Sigma}V^*)^*y \tag{14}$$

$$V\hat{\Sigma}\hat{U}^*\hat{U}\hat{\Sigma}V^*c = V\hat{\Sigma}\hat{U}^*y. \tag{15}$$

V et $\hat{\Sigma}$ sont inversibles, et $U^*U = I$, donc

$$\hat{\Sigma}V^*c = \hat{U}^*y. \tag{16}$$

L'algorithme se déroule ainsi:

Algorithme 1. Régression linéaire.

1. *Calculer une SVD réduite:* $A = \hat{U}\hat{\Sigma}V^*$.
2. *Calculer le vecteur* $\hat{U}^*y$.
3. *Résoudre le système diagonal* $\hat{\Sigma}z = \hat{U}^*y$ *pour* z.
4. *Ecrire la solution* $c = Vz$.

Exercice 1. *Régression linéaire.* Ecrivez en Python une routine pour faire une régression linéaire sur deux vecteurs $x, y \in \mathbb{R}^m$. Dans une régression linéaire, on cherche $a, b \in \mathbb{R}$ tels que $y \sim ax + b$ dans le sens des moindres carrés, c.-à-d. qu'on cherche un vecteur $c = {}^t(a, b)$ tels que la norme $||y - (ax + b)||$ est minimisée. Pour ce faire, écrire le problème comme un problème d'ajustement polynomial (de degré 1) sous la forme $Ac = y$ et trouver c qui minimise le vecteur de résidus $r = y - Ac$ à l'aide de la décomposition en valeurs singulières et de l'algorithme présenté. Générez des vecteurs x et y et testez votre routine.

Exercice 2. Vérifiez que la matrice UU^* est un projecteur orthogonal sur im A.

References

[1] Jean Fresnel. *Algebre des matrices*. Hermann, 2013.

[2] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.