Régression Linéaire Multiple

Benoit Gaüzère, Stéphane Canu
benoit.gauzere@insa-rouen.fr

INSA Rouen Normandie - ITI

April 5, 2023

Rappels sur la régression linéaire

Problème

x : Hauteur (en m) des arbres dominants

y : Logarithme du nombre de nids de processionnaires par arbre

Modèle

$$y = ax + b + \varepsilon$$

on cherche à expliquer la variable y par la variable x

Résolution

On dispose d'un ensemble d'observations:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i), \ldots, (x_n, y_n),$$

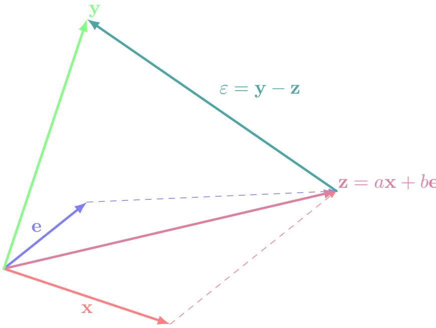
et on cherche à estimer a et b à partir de ces données.

Version vectorielle dans $\mathbb{R}^n$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Écritures Matricielles

- Moyenne : $\mathbf{x}^\top \mathbf{e} = \sum x_i = n\,\overline{x}$, $\mathbf{y}^\top \mathbf{e} = \sum y_i = n\,\overline{y}$
- Variance/covariance :
  - $||\mathbf{x} - \overline{x}\,\mathbf{e}||^2 = n\,V(\mathbf{x})$
  - $||\mathbf{y} - \overline{y}\,\mathbf{e}||^2 = n\,V(\mathbf{y})$
  - $(\mathbf{x} - \overline{x}\,\mathbf{e})^\top (\mathbf{y} - \overline{y}\,\mathbf{e}) = n\,\mathrm{cov}(\mathbf{x}, \mathbf{y})$
- Régression : pour tout couple (a, b)
  - $\mathbf{z} = a\mathbf{x} + b\mathbf{e}$
  - $\varepsilon = \mathbf{y} - \underbrace{a\mathbf{x} + b\mathbf{e}}_{\mathbf{z}}$



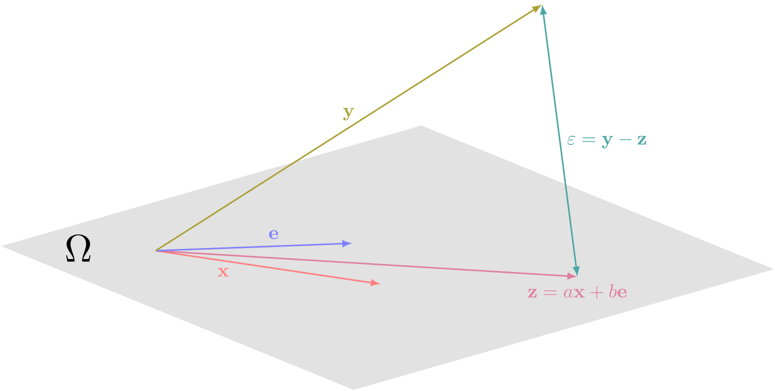
L'Ensemble des Modèles

Un modèle $\Leftrightarrow$ un couple (a, b) tel que $\left\{\begin{array}{ll} y_1 & = ax_1 + b + \varepsilon_1 \\ \vdots & \qquad \vdots \\ y_i & = ax_i + b + \varepsilon_i \\ \vdots & \qquad \vdots \\ y_n & = ax_n + b + \varepsilon_n \end{array}\right.$

$\mathbf{x}, \mathbf{e}, \varepsilon$, $\mathbf{y}$ et $\mathbf{z} \in \mathbb{R}^n$, et on a $\mathbf{y} = \underbrace{a\mathbf{x} + b\mathbf{e}}_{\mathbf{z}} + \varepsilon$

Plan Ω engendré par $\mathbf{x}$ et $\mathbf{e}$

- $\Omega = \left\{\mathbf{z} \in \mathbb{R}^n | \exists (a,b) \in \mathbb{R}^2, \mathbf{z} = a\mathbf{x} + b\mathbf{e}\right\}$
- $\mathbf{x}, \mathbf{e}$ et $\mathbf{z} \in \Omega$



Quel est le meilleur modèle?

Représentation Matricielle

Les Données

- Matrice des observations : $X = [\mathbf{x} \quad \mathbf{e}]$
- Inconnues : $\boldsymbol{\alpha} = [a \quad b]$
- Variable à expliquer : $\mathbf{y} \in \mathbb{R}^n$

Le Modèle

$$X\boldsymbol{\alpha} + \varepsilon = \mathbf{y}$$

$$\begin{array}{cc} & \begin{bmatrix} a \\ b \end{bmatrix} \\ \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_i & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} & \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix} \end{array}$$

Version matricielle de la Fonction de Coût

Moindres Carrés

Le problème des moindres carrés peut se réécrire de la manière suivante :

$$\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) \qquad \text{avec} \qquad J(\boldsymbol{\alpha}) = \frac{1}{2}||\mathbf{y} - X\boldsymbol{\alpha}||^2$$

$$\begin{aligned}
\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) &= \tfrac{1}{2}||\mathbf{y} - X\boldsymbol{\alpha}||^2 \\
&= \tfrac{1}{2}(\mathbf{y} - X\boldsymbol{\alpha})^\top(\mathbf{y} - X\boldsymbol{\alpha}) \\
&= \tfrac{1}{2}\mathbf{y}^\top\mathbf{y} - \tfrac{1}{2}\boldsymbol{\alpha}^\top X^\top\mathbf{y} - \tfrac{1}{2}\mathbf{y}^\top X\boldsymbol{\alpha} + \tfrac{1}{2}\boldsymbol{\alpha}^\top X^\top X\boldsymbol{\alpha} \\
&= \tfrac{1}{2}\mathbf{y}^\top\mathbf{y} - \boldsymbol{\alpha}^\top X^\top\mathbf{y} + \tfrac{1}{2}\boldsymbol{\alpha}^\top X^\top X\boldsymbol{\alpha}
\end{aligned}$$

NB : $\mathbf{y}^\top\mathbf{y}$ est un scalaire, $X^\top\mathbf{y}$ est un vecteur de $\mathbb{R}^2$ et $X^\top X$ est une matrice 2×2.

Dérivées partielles du coût

$$J(\boldsymbol{\alpha}) \quad = \quad \frac{1}{2}\mathbf{y}^\top\mathbf{y} \quad - \quad \boldsymbol{\alpha}^\top X^\top \mathbf{y} \quad + \quad \frac{1}{2}\boldsymbol{\alpha}^\top X^\top X\boldsymbol{\alpha}$$

$$\frac{\partial J(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \quad = \quad 0 \quad - \quad X^\top\mathbf{y} \quad + \quad X^\top X\boldsymbol{\alpha}$$

Démonstration par éléments de $\boldsymbol{\alpha}$

- Soit $\mathbf{w} = X^\top\mathbf{y}$ et $M = X^\top X$
- $\dfrac{\partial \boldsymbol{\alpha}^\top\mathbf{w}}{\partial \alpha_i} = \dfrac{\partial \sum_{j=1}^{p+1} w_j\alpha_j}{\partial \alpha_i} = w_i$
- $\dfrac{\partial \boldsymbol{\alpha}^\top M\boldsymbol{\alpha}}{\partial \alpha_i} = \dfrac{\partial \sum_{j=1}^{p+1}\sum_{k=1}^{p+1} \alpha_j\alpha_k M_{jk}}{\partial \alpha_i} = \displaystyle\sum_{j=1}^{p+1}\alpha_j M_{ji} + \sum_{k=1}^{p+1}\alpha_k M_{ik}$,

  car $(uv)' = uv' + u'v$ avec $u = \alpha_j$ et $v = \sum_{k=1}^{p+1}\alpha_k M_{jk}$

$$\nabla J(\boldsymbol{\alpha}) = -\mathbf{w} + M\boldsymbol{\alpha} = -X^\top\mathbf{y} \; + \; X^\top X\boldsymbol{\alpha}$$

La Minimisation du Coût

Méthode du Gradient

La minimisation du cout $J(\boldsymbol{\alpha}^\star)$ est réalisée lorsque le gradient du cout s'annule, soit lorsque :

$$\nabla J(\boldsymbol{\alpha}^\star) = 0 \Leftrightarrow -X^\top \mathbf{y} \;+\; X^\top X \boldsymbol{\alpha}^\star = 0$$

Solution

La solution du problème de minimisation des moindres carrés est le vecteur $\boldsymbol{\alpha}^\star$ défini par :

$$\boldsymbol{\alpha}^\star = \left(X^\top X\right)^{-1} X^\top \mathbf{y}$$

Prédiction

$\mathbf{z} = H\mathbf{y}$

$$\begin{aligned} \mathbf{z} &= X\boldsymbol{\alpha}^\star \\ &= X\left(X^\top X\right)^{-1} X^\top \mathbf{y} \\ &= H\mathbf{y} \end{aligned}$$

avec $H = X\left(X^\top X\right)^{-1} X^\top$

Les moindres carrés comme une projection

- H est un projecteur de $\mathbb{R}^n$
- $HH = H$

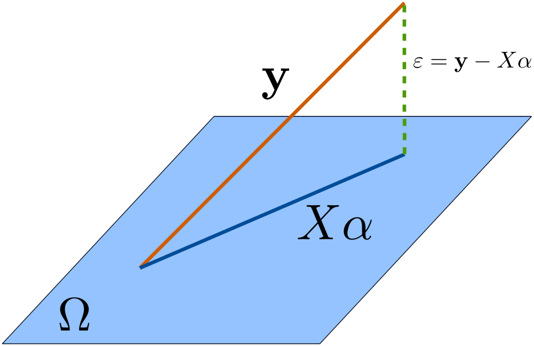
$$\begin{aligned} HH &= X\left(X^\top X\right)^{-1} X^\top \quad X\left(X^\top X\right)^{-1} X^\top \\ &= X\left(X^\top X\right)^{-1} \left(X^\top X\left(X^\top X\right)^{-1}\right) X^\top \\ &= X\left(X^\top X\right)^{-1} X^\top \qquad = H \end{aligned}$$

La régression du point de vue géométrique

Interprétation

- $\Omega = \mathrm{span}\{\mathbf{x}, \mathbf{e}\}$; $\mathbf{z} \in \Omega \Leftrightarrow \exists\, a, b \in \mathbb{R} \quad \mathbf{z} = a\mathbf{x} + b\mathbf{e}$
- Régression: projection du vecteur $\mathbf{y}$ sur Ω par H:

$$X\boldsymbol{\alpha}^\star = H\mathbf{y} \qquad \text{avec} \qquad H = X\left(X^\top X\right)^{-1}X^\top \tag{1}$$

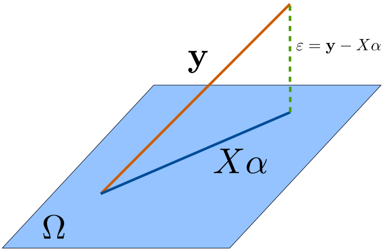


Estimation et orthogonalité

Minimisation géométrique

- $J(\boldsymbol{\alpha}) = \dfrac{1}{2} \|\varepsilon\|^2$
- $\|\varepsilon\|^2$ est minimal quand on choisi $\boldsymbol{\alpha}$ tel que $\mathbf{z} = X\boldsymbol{\alpha}$ est la projection orthogonale de $\mathbf{y}$ sur Ω :

$$\begin{aligned} X^\top(\mathbf{y} - X\boldsymbol{\alpha}^\star) = 0 \quad & \Leftrightarrow \quad X^\top\mathbf{y} - X^\top X\boldsymbol{\alpha}^\star = 0 \\ & \Leftrightarrow \quad X^\top X\boldsymbol{\alpha}^\star = X^\top\mathbf{y} \\ & \Leftrightarrow \quad \boldsymbol{\alpha}^\star = (X^\top X)^{-1}X^\top\mathbf{y} \end{aligned}$$



La Régression multiple

Comment étendre la régression à plusieurs variables explicatives ?

La Régression multiple

- Variables explicatives :
  1. altitude (en m)
  2. pente (en degré)
  3. nombre de pins moyens dans une placette de 5 ares
  4. hauteur de l'arbre échantillonné au centre de la placette
  5. diamètre de cet arbre
  6. note de densité du peuplement
  7. orientation de la placette (1 : sud, 2 : autres)
  8. hauteur (en m) des arbres dominants
  9. nombre de strates de végétation
  10. mélange du peuplement (1 : pas mélangé, 2 : mélangé)
- Variable à expliquer (y) :
  - logarithme du nombre de nids de processionnaires par arbre d'une placette

source : Tomassone et al. [1992]

Objectifs de la Régression Multiple

p coefficients à estimer

$$
\begin{aligned}
y =& a_0 + a_1 x_1 + a_2 x_2 + \dots a_j x_j + \dots + a_p x_p + \varepsilon \\
y =& a_0 + \sum_{j=1}^{p} a_j x_j + \varepsilon
\end{aligned}
$$

NB: si $p = 1$ c'est la régression linéaire simple

Comment faire ?

$\Rightarrow$ Adaptons la régression linéaire simple sous forme matricielle à p colonnes

Représentation des Données

Adaptation d'$\boldsymbol{\alpha}$

- Un coefficient a_j pour chaque variable explicative
- Plus a_0 pour le décalage à l'origine
- $\boldsymbol{\alpha} \in \mathbb{R}^{p+1}$, avec $\boldsymbol{\alpha}(j) = a_j$:

$$\boldsymbol{\alpha}^\top = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_j & \dots & a_p \end{pmatrix}$$

Matrice des données

- $(\mathbf{x}_i, y_i) \in \mathbb{R}^p \times \mathbb{R}, i = 1, \dots, n$
- $X \in \mathbb{R}^{n \times p+1}$, avec $X(i,:) = (1, \mathbf{x}_i^\top)$

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

La Régression Multiple : le Modèle Matriciel

$$\mathbf{y} = X\boldsymbol{\alpha} + \varepsilon$$

Résolution

- Méthode géométrique ou du gradient
- Hypothèse: X est une matrice de rang $p+1$ et donc la matrice $X^\top X$ est inversible
- $\boldsymbol{\alpha}^\star = (X^\top X)^{-1} X^\top \mathbf{y}$
- On résoud un système linéaire de $p+1$ inconnues et $p+1$ équations

Mise en œuvre

$$\boldsymbol{\alpha}^\star = (X^\top X)^{-1} X^\top \mathbf{y}$$

Comment résoudre ce système linéaire ?

- Méthode directe
  1. Calcul de l'inverse : $M = \mathsf{inv}(X^\top X)$
  2. Calcul des alpha : $\boldsymbol{\alpha}^\star = M(X^\top \mathbf{y})$
- Factorisation de Cholesky
  - $\boldsymbol{\alpha}^\star = (X^\top X) \backslash (X^\top \mathbf{y})$
- Factorisation QR
  - $X = QR$, Q matrice orthogonale, et R, une matrice triangulaire supérieure
  - $$\begin{aligned}(R^\top Q^\top Q R)\boldsymbol{\alpha}^\star =& R^\top Q^\top \mathbf{y} \\ R\boldsymbol{\alpha}^\star =& Q^\top \mathbf{y}\end{aligned}$$

La régression multiple en Python

Calcul des estimateurs

- Attention à a_0
- Pas besoin d'inverser la matrice

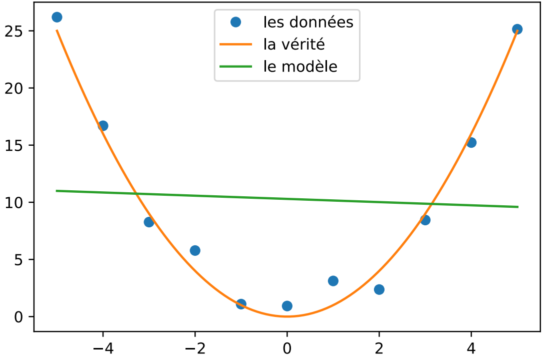
```
def regression_fit(X,y):
    n,p = X.shape
    X = np.concatenate([np.ones((n,1)), X], axis=1)
    alpha = np.linalg.solve(X.T@X, X.T@y)
    return alpha
```

Calcul de la prédiction

```
def regression_predict(X,alpha):
    return X@alpha[1:] + alpha[0]
```

Extension aux problèmes non-linéaires

Que faire dans le modèle n'est pas adapté ?



Extension aux problèmes non-linéaires

Projection des données

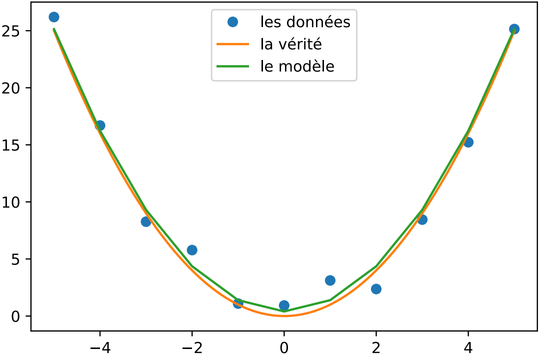
- Projeter les données dans un espace où la relation est linéaire
- $\Phi : \mathbb{R}^p \to \mathbb{R}^d$
- par ex.:
  - $\Phi(x) = x^2$
  - $\Phi_c(x) = \begin{cases} 0 \text{ si } x < c \\ x - c \text{ si } x \geq c \end{cases}$

Résolution dans le nouvel espace

- On applique la régression dans le nouvel espace
- Mais ...
  - Comment connaître la projection ?
  - Que se passe t-il si $d > n$?

Extension au non-linéaire

$$\Phi(x) = x^2$$



Extension à d'autres types de problèmes

$$\mathbf{y} = X\mathbf{a} + \varepsilon$$

Variable à expliquer $\mathbf{y}$	Variable explicative X	Méthode
Quantitative	Quantitative	Régression
Quantitative	Qualitative	Analyse de la Variance
Qualitative	Quantitative	Régression Logistique
Qualitative	Qualitative	Analyse Discriminante

Exemples

- Variété de maïs et rendement à l'hectare
- Longueur des pétales et espèces de plante
- Présence de mots et Spam
- ...

Conclusion

Le Modèle Linéaire

$\mathbf{y} = X\boldsymbol{\alpha} + \varepsilon$

Optimisation au sens des moindres carrés

- $\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\varepsilon\|^2$
- Le cout est minimal quand $X^\top(\mathbf{y} - X\boldsymbol{\alpha}^\star) = 0$
- Calcul de $\boldsymbol{\alpha}^\star$ au prix de la résolution d'un système linéaire

Reste à faire

- Diagnostic du modèle
- Diagnostic des variables
- Diagnostic des individus

Repéres bibliographiques

- https://fr.wikipedia.org/wiki/R%C3%A9gression_lin%C3%A9aire_multiple
- https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/14/lecture-b*14.pdf
- https://www.coursera.org/lecture/erasmus-b*econometrics/lecture-b*2-b*2-b*on-b*multiple-b*regression-b*represent
- https://web.stanford.edu/class/stats202/slides/Linear-b*regression.html