




ANALYSE STATISTIQUE & LANGAGE R

Modélisation avancée
Séance complémentaire



Déroulement

- Les modèles linéaires généralisés (GLM)
 - *Linéaire* $]-\infty ; \infty[$
 - Distribution normale
 - *Logistique (binomiale)* $[0 ; 1]$
 - Distribution binomiale
 - *Poisson* $[0 ; \infty[$
 - Distribution de poisson
 - *Multinomiale / Polynomiale...*
- Méthodes de résolution
 - *OLS, 2OLS, méthode des moindres carrés, doubles moindres carrés...*
 - *SGD, descente de gradient*
- Régressions pénalisées
 - *Ridge*
 - *Lasso*
 - *Elastic Net*

Les modèles GLM

- Les GLM sont une **extension** des modèles linéaires classiques, qui peuvent être utilisés lorsque les réponses ne sont pas de type **numérique continues**.
- Les GLM sont utilisés dans deux situations :
 - *Lorsque les données sont de **type comptage** (nombre d'évènements dans un laps de temps : nombre de visites sur un site, défaut de crédit, nombre de crises/récidives après un traitement médical, accidents...)*
 - *Lorsque les données sont de **type binaire** (0/1)*

Fonction de lien

Type des réponses (et des erreurs)	Domaine de définition des réponses	Distribution des erreurs (et des réponses)	Nom de la fonction de lien	Fonction de Lien	Fonction de la moyenne	Fonction de la variance
Quantitatif continu	Réel] $-\infty$; $+\infty$ [Gaussienne	Identité	$\sum_{j=1}^p x_{ij} \beta_j = \mu$	$\mu = \sum_{j=1}^p x_{ij} \beta_j$	$var(y_i) = cste$
Comptage	Entier [0 ; $+\infty$ [Poisson	Log	$\sum_{j=1}^p x_{ij} \beta_j = \ln(\mu)$	$\mu = \exp\left(\sum_{j=1}^p x_{ij} \beta_j\right)$	$var(y_i) = \mu$
Binaire (oui/non)	Entier [0 ; 1]	Binomiale	Logit	$\sum_{j=1}^p x_{ij} \beta_j = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{1}{1 + \exp\left(-\sum_{j=1}^p x_{ij} \beta_j\right)}$	$var(y_i) = \mu \frac{(n-\mu)}{n}$

Les données sont mises à l'échelle pour être modélisables, via une relation linéarisée

Pour retrouver les valeurs d'origines, il suffit d'appliquer la fonction inverse : **ln()** => **exp()**

Exemple de régression de poisson

- <https://www.kaggle.com/gauravduttakiit/award-competition>
- Déterminer le nombre de récompenses mathématique vis-à-vis du score de chaque individu

Régressions pénalisées

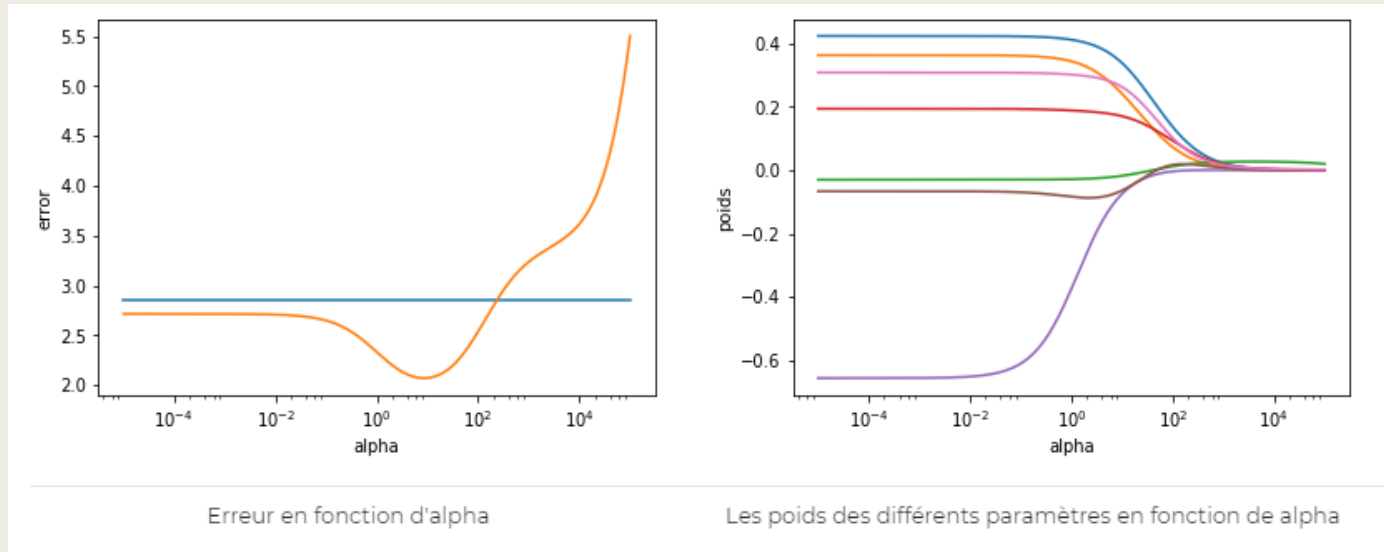
- Les régressions pénalisées sont utilisables pour les différents types de régressions : linéaire gaussien, poisson, binomiale.
- Cela permet une sélection de variables, mais aussi une partie de régularisation, c'est à dire qu'une régression trop complexe/spécifique sera pénalisée, au profit d'une régression plus générique
 - *Réduction de certains poids / coefficients (jusqu'à 0 pour exclure la variable)*
 - *Rétrécir la plage de données que peuvent prendre les variables*
 - *Pénaliser les coefficients ayant des degrés élevés...*

Lasso et Ridge

L1 Regularization or Lasso	L2 Regularization or Ridge
Dispose de sélection de variables Les poids peuvent être mis à 0	Les poids sont pénalisés, mais non mis à 0
Robuste aux valeurs extrêmes	Non robuste aux valeurs extrêmes
Peu efficace avec un grand nombre de dimensions	Capable d'apprendre sur de complexes patterns de données
Choisi une variable dans un groupe de variables corrélés (souvent celle la plus liée à la cible), et met les autres à 0	Partage les poids entre plusieurs variables corrélées Les variables initiales se retrouvent donc dans le modèle finale (on ne sait pas lesquelles sont les plus importantes)

Lasso, ridge, elastic net

- L1 = lasso / L2 = ridge
 - $\text{Alpha} = 0 \Rightarrow \text{ridge}$
 - $\text{Alpha} = 1 \Rightarrow \text{lasso}$
 - Valeur entre 0 et 1 $\Rightarrow \text{Elastic net (mix des deux)}$
- Lambda : intensité de la régression pénalisée, plus la valeur est importante, plus la régularisation l'est aussi



* $\text{Lambda} \Rightarrow \text{Alpha}$
 $\text{Alpha} \Rightarrow \text{l1_ratio}$

Gradient descent

- La méthode de descente de gradient, peut également être appliquée aux régressions.
 - **Descente de Gradient classique** : On fait passer la totalité des observations, le gradient est calculé, les coefficients sont corrigés...
 - **Online / SGD (Stochastic descent gradient)** : Gradient calculé pour chaque observation, correction des coefficients...
 - **Mini-batch** (traitement par lots) : On fait passer des lots d'observations. Calcul du gradient pour chaque lot, correction des coefficients...
- ⇒ Le traitement par lots permet d'améliorer la convergence en réduisant le nombre de passage sur la base entière
- ⇒ Il permet également de se contenter **de charger partiellement les données** en mémoire au fur et à mesure
- ⇒ Eviter des lots trop petits, pour ne pas stagner sur un minimum local

Grad. Des. – Taux d'apprentissage

■ Le taux d'apprentissage

Fait varier les poids/coefficients plus ou moins vite pour converger.

La stratégie peut être d'avoir un taux d'A. qui décroît au fur et à mesure des itérations, afin d'avancer rapidement au départ, et d'ajuster doucement ensuite (ex : le golf).

