

Régularisation en régression multivariable

1 Cadre général : modèle multivariable

On considère un problème de régression avec plusieurs variables explicatives :

$$x = (x_1, x_2, \dots, x_p)$$

Le modèle linéaire multivariable s'écrit :

$$\hat{y} = w_0 + \sum_{j=1}^p w_j x_j = w_0 + Xw$$

La fonction de perte quadratique moyenne (MSE) est :

$$\mathcal{L}(w) = \frac{1}{n} \sum_{i=1}^n (y_i - x_i^\top w)^2$$

Lorsque le nombre de variables p est élevé, ou que certaines variables sont corrélées, le modèle peut devenir instable et sur-apprendre les données.

2 Principe de la régularisation

La régularisation consiste à ajouter une pénalité sur les coefficients du modèle afin de limiter sa complexité :

$$\min_w \mathcal{L}(w) + \lambda \Omega(w)$$

- λ contrôle la force de la régularisation
- $\Omega(w)$ est le terme de pénalisation

3 Régularisation L2 (Ridge)

3.1 Formulation mathématique

$$\min_w \frac{1}{n} \|y - Xw\|_2^2 + \lambda \|w\|_2^2$$

3.2 But de la régularisation L2

Le but principal de la régularisation L2 est de **stabiliser le modèle multivariable** en réduisant la variance des coefficients.

Elle est particulièrement efficace lorsque :

- les variables explicatives sont fortement corrélées
- le nombre de variables est élevé
- le problème est mal conditionné

3.3 Effet sur les coefficients

- Tous les coefficients sont réduits de manière continue
 - Aucun coefficient n'est exactement nul
 - Les variables corrélées partagent l'information
- La solution analytique devient :

$$\hat{w} = (X^\top X + \lambda I)^{-1} X^\top y$$

4 Régularisation L1 (Lasso)

4.1 Formulation mathématique

$$\min_w \frac{1}{n} \|y - Xw\|_2^2 + \lambda \sum_{j=1}^p |w_j|$$

4.2 But de la régularisation L1

Le but principal de la régularisation L1 est la **sélection automatique de variables** dans un contexte multivariable.

Elle est utilisée lorsque :

- beaucoup de variables sont non pertinentes
- l'interprétabilité du modèle est importante
- on souhaite un modèle parcimonieux (sparse)

4.3 Effet sur les coefficients

- De nombreux coefficients deviennent exactement nuls
- Le modèle ne conserve qu'un sous-ensemble de variables
- En présence de variables corrélées, une seule variable est généralement sélectionnée

5 Régularisation Elastic Net

5.1 Formulation mathématique

$$\min_w \frac{1}{n} \|y - Xw\|_2^2 + \lambda (\alpha \|w\|_1 + (1 - \alpha) \|w\|_2^2)$$

avec $\alpha \in [0, 1]$.

5.2 But de l'Elastic Net

L'Elastic Net vise à **combinaison des avantages de L1 et L2** :

- sélection de variables (L1)
- stabilité numérique (L2)

Il est particulièrement adapté aux problèmes multivariables réels où les variables sont corrélées.

5.3 Effet sur les coefficients

- Certains coefficients sont nuls
- Les variables corrélées peuvent être sélectionnées ensemble
- Le modèle est plus stable que le Lasso seul

6 Comparaison des objectifs en multivariable

Méthode	But principal	Effet sur les coefficients	Corrélation
L2 (Ridge)	Réduction de la variance	Poids réduits	Très bon
L1 (Lasso)	Sélection de variables	Poids nuls	Faible
Elastic Net	Compromis	Zéros + stabilité	Bon

7 Impact sur la généralisation

- Sans régularisation : variance élevée
- L2 : légère augmentation du biais, forte réduction de la variance
- L1 : modèle simple, interprétable mais plus biaisé
- Elastic Net : équilibre biais / variance optimal

8 Conclusion

Dans un contexte multivariable :

- la régularisation est essentielle pour éviter l'instabilité
- L2 est privilégiée pour la prédiction
- L1 est privilégiée pour la sélection de variables
- Elastic Net offre le meilleur compromis dans la majorité des cas