

Régularisation L1 / L2 et Algorithmes Proximaux

Chapitre 4 – Optimisation non lisse

Fatimetou Limam Abeid
C16698

Université FST

2025–2026

Plan

- 1 Introduction
- 2 Régularisation L2 et L1
- 3 Algorithmes Proximaux et ISTA
- 4 TP2 : Sélection de variables
- 5 Questions de Synthèse
- 6 Conclusion

Introduction

- En apprentissage automatique, les modèles complexes peuvent sur-apprendre.
- La régularisation permet :
 - d'améliorer la généralisation
 - de stabiliser les solutions
 - de rendre les modèles interprétables
- Le chapitre 4 étudie :
 - la régularisation L2 (Ridge)
 - la régularisation L1 (Lasso)
 - les algorithmes proximaux (ISTA/FISTA)

Régularisation L2 (Ridge)

$$\min_w f(w) + \lambda \|w\|_2^2$$

- Pénalise les grands coefficients
- Stabilise les solutions
- **Ne produit pas de zéros exacts**

Régularisation L1 (Lasso)

$$\min_w f(w) + \lambda \|w\|_1$$

- Favorise la parcimonie (sparsity)
- Produit des coefficients exactement nuls
- Réalise une sélection automatique de variables

Interprétation géométrique

- La boule L₂ est ronde → peu de solutions sur les axes
- La boule L₁ est anguleuse → solutions sur les axes
- ⇒ Apparition naturelle de coefficients nuls

Opérateur Proximal

$$\text{prox}_{\gamma\Omega}(v) = \arg \min_u \left(\Omega(u) + \frac{1}{2\gamma} \|u - v\|^2 \right)$$

- Généralise la projection
- Permet de traiter les fonctions non lisses

Soft-Thresholding (Norme L1)

$$\text{prox}_{\gamma \|\cdot\|_1}(v) = \text{sign}(v) \max(0, |v| - \gamma)$$

- Annule les petites valeurs
- Conserve les grandes valeurs
- Favorise la parcimonie

Algorithme ISTA

Problème du Lasso :

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

Itération ISTA :

$$w^{k+1} = \text{prox}_{\lambda/L} \left(w^k - \frac{1}{L} \nabla f(w^k) \right)$$

- Gradient sur la partie lisse
- Seuillage doux sur la partie non lisse

Dataset Breast Cancer

- Comparaison Ridge vs Lasso
- Normalisation des données
- Analyse du chemin du Lasso

Résultats observés

- Ridge : coefficients petits mais non nuls
- Lasso : nombreuses variables éliminées
- Meilleure interprétabilité avec L1

Question 1 : L1 vs L2

Pourquoi la norme L1 est-elle plus adaptée à la sélection de variables ?

- La norme L1 produit des coefficients exactement nuls
- La norme L2 ne fait que réduire leur amplitude
- La géométrie de L1 favorise les solutions sur les axes

Conclusion : L1 permet une sélection automatique de variables.

Question 2 : Rôle de λ dans ISTA

Effet du paramètre λ :

- λ petit : image proche de l'image bruitée
- λ moyen : bon compromis bruit / détails
- λ grand : image trop lissée

λ contrôle le compromis :

Fidélité aux données \leftrightarrow Parcimonie

Conclusion

- La régularisation L1 favorise la parcimonie
- Les algorithmes proximaux rendent l'optimisation non lisse efficace
- Le paramètre λ est crucial pour la qualité du modèle

Méthodes clés pour des modèles interprétables