



JOURNÉE DATA SCIENCE

Mélanie GOMIS
Arnaud MILLET
Enzo NGUYEN
Cédric RANDRIAMAROTIA
Maximiliano RUIZ



SOMMAIRE

1. Introduction

2. Création des données

3. Différents tests

4. Résultats

5. Conclusion

INTRODUCTION

Quels sont les meilleurs tests à faire pour sélectionner les différentes variables explicatives pour une régression linéaire ?



CREATION DES DONNÉES

- fonction generate_lm
 - Des données X sans corrélation entre elles
- fonction.lm.longitudinal
 - Des données X avec corrélation longitudinales

DIFFERENTS TESTS

- Student
- Recherche Exhaustive
- Stepwise

▲ Student
X1 0.0000000
X2 0.0000000
X3 -1.1610531
X4 0.0000000
X5 -0.1932012
X6 0.0000000
X7 0.2691992
X8 0.0000000
X9 0.0000000
X10 0.0000000

RÉSULTATS

comparaison	selection	Nombre de meilleurs modèles
Erreur moyenne de prédition calculée sur l'ensemble test	Algorithme de sélection pas à pas (stepwise)	7
Erreur moyenne de prédition calculée sur l'ensemble test	Recherche exhaustive (Best subset avec l'algorithme Leaps and bound)	11
Erreur moyenne de prédition calculée sur l'ensemble test	Test de Student de signification des coefficients	12
Erreur quadratique moyenne de β^{\wedge} (RMSE) sur l'ensemble train	Algorithme de sélection pas à pas (stepwise)	7
Erreur quadratique moyenne de β^{\wedge} (RMSE) sur l'ensemble train	Recherche exhaustive (Best subset avec l'algorithme Leaps and bound)	11
Erreur quadratique moyenne de β^{\wedge} (RMSE) sur l'ensemble train	Test de Student de signification des coefficients	12
Sensibilité du support estimé	Algorithme de sélection pas à pas (stepwise)	16
Sensibilité du support estimé	Recherche exhaustive (Best subset avec l'algorithme Leaps and bound)	15
Sensibilité du support estimé	Test de Student de signification des coefficients	15

On a fait varier : n, le bruit et la corrélation longitudinale.

CONCLUSION

- Difficultés rencontrés :
 - À propos des données
 - La gestion du temps
 - Les résultats
- Perspectives :
 - Tableau des résultats
 - Autres types de données

MERCI POUR
VOTRE ATTENTION

