



Analyse multivariée



ECTS
5 crédits



Composante
Faculté des
Sciences

Présentation

Description

La taille des données statistique ne cesse de croître, et notamment la richesse de la description des unités statistiques. Or, la modélisation statistique linéaire classique devient invalide en grande dimension, c'est-à-dire lorsque le nombre des variables dépasse celui des unités statistiques. Ce cours présente les techniques les plus courantes utilisées pour régulariser les modèles linéaires en grande dimension.

Objectifs

Former à la modélisation linéaire uni- et multi-variée en grande dimension, c'est-à-dire à diverses techniques de régularisation de la modélisation linéaire classique.

Pré-requis nécessaires

Cours d'analyse des données multidimensionnelle (ACP & CA). Cours de géométrie euclidienne, d'espaces vectoriels normés et de réduction des endomorphismes.

Pré-requis recommandés : Cours de statistique descriptive univariée et bivariée. Bonne maîtrise du calcul matriciel.

Syllabus

Introduction

Données de grande dimension. Réduction dimensionnelle et régularisation.

I - Modélisation linéaire régularisée d'une variable continue.

1. Le modèle linéaire classique.

- a) Rappels express.
- b) Les pannes dues aux colinéarités.

2. Régression sur composantes principales.

- a) La méthode.
- b) Qualités et défauts.

3. Régression PLS.

- a) Critère et programme de rang 1.
- b) Critère et programme de rangs ultérieurs.
- c) Pourquoi PLS régularise.
- d) Choix du nombre de composantes pour la prédiction.
- e) Métrique du continuum entre OLS et PLS.

4. Régressions linéaires pénalisées.

- a) Régression ridge.



b) LASSO.

c) Elastic net.

II - Modélisation linéaire régularisée d'un groupe de variables continues.

1. Le modèle linéaire gaussien multivarié

a) Le modèle classique.

b) Le modèle pénalisé.

c) Le modèle de MANOVA.

2. Régression PLS multivariée.

a) Critère et programme de rang 1 avec métriques quelconques.

b) Cas particuliers: analyse canonique, ACP sur Variables Instrumentales, Régression PLS2.

c) Critère et programme de rangs ultérieurs.

d) Prédiction: choix du nombre optimal de composantes.

e) Métriques du continuum entre Analyse Canonique, ACPVI et PLS.

III - Modélisation linéaire d'une variable nominale: analyses discriminantes linéaires.

1. Analyse factorielle discriminante

a) Critère et programme.

b) Composantes et pouvoirs discriminants.

2. Analyse discriminante PLS.

a) Critère et programme.

b) Composantes et pouvoirs discriminants.

c) Analyse discriminante barycentrique.

3. Aspects décisionnels.

a) Décision (classement), pertes, règles de décision (affectation), risques.

b) Choix du bon nombre de composantes pour la décision.

Informations complémentaires

Volumes horaires :

CM : 21

TD :

TP :

Terrain :

Infos pratiques

Contacts

Responsable pédagogique

Elodie Brunel-piccinini

+33 4 67 14 41 64

elodie.brunel-piccinini@umontpellier.fr

Responsable pédagogique

Xavier Bry

+33 4 67 14 35 78

xavier.bry@umontpellier.fr