



HAL
open science

Sélection de modèle pour la classification non supervisée

Jean-Patrick Baudry

► **To cite this version:**

Jean-Patrick Baudry. Sélection de modèle pour la classification non supervisée. Journées MAS et Journée en l'honneur de Jacques Neveu, Aug 2010, Talence, France. inria-00510149

HAL Id: inria-00510149

<https://inria.hal.science/inria-00510149>

Submitted on 17 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journées MAS 2010, Bordeaux

Session : Apprentissage statistique

Sélection de modèle pour la classification non supervisée

par **Jean-Patrick Baudry**

Nous rappelons les bases de l'approche de la classification non supervisée par les modèles de mélange. La méthode usuelle repose sur le maximum de vraisemblance et le choix du nombre de classes à former se fait par des critères pénalisés. Nous nous intéressons particulièrement au critère ICL (Biernacki, Celeux et Govaert, 2000), mis au point pour tenir compte de l'objectif de classification et pertinent en pratique. L'étude que nous proposons de ce critère et de la notion de classe sous-jacente repose sur l'introduction d'un cadre de minimisation d'un contraste adapté à ce contexte. Ce faisant nous définissons un nouvel estimateur et une nouvelle famille de critères de sélection de modèles dont nous étudions les propriétés – notamment la consistance. La calibration de ces critères peut se faire par l'heuristique de pente (Birgé et Massart, 2006). Divers aspects pratiques de leur mise en œuvre sont discutés et leur comportement pratique illustré par des simulations.

Adresse :

Jean-Patrick BAUDRY

Laboratoire de Mathématiques, Université Paris Sud

Bâtiment 430, bureau 25

91405 Orsay Cedex, France

E-mail : jean-patrick.baudry@math.u-psud.fr

<www.math.u-psud.fr/~baudry>

Session : Apprentissage statistique