

Cours de Statistique approfondie 2004-2005

Projet

1. Cas d'un article

Le travail personnel, de l'ordre de 15 pages dactylographiées, consiste à analyser un article scientifique : il ne s'agit pas de le traduire ni de le paraphraser mais d'en faire un exposé synthétique destiné à un statisticien qui ne l'a pas lu et veut savoir de quoi il retourne. Certains articles peuvent nécessiter une recherche de documents complémentaires. Une application sur des données personnelles des méthodes exposées dans l'article sera grandement appréciée. Lorsque les notations utilisées diffèrent trop de celles du cours, il sera apprécié de les reformuler.

2. Cas d'une étude de logiciel

Le travail pourra être fait à deux. Il consistera en une présentation de la méthode et du programme (ou de la procédure analysée) avec un « manuel » simple d'utilisation et une application à des données réelles .

Procédure de choix

Chaque élève (ou binôme éventuellement pour les logiciels) régulièrement inscrit au cours propose par courrier électronique 3 sujets classés par ordre de préférence. Les premiers arrivés sont les premiers servis avec priorité pour les élèves qui assistent au cours (!). La liste est mise à jour au fur et à mesure : les sujets surlignés en jaune ne peuvent plus être choisis. Si les 3 choix sont déjà pris, on recommence.

Remise du projet

La note de lecture, ou l'étude de logiciel, devra être remise au plus tard le 15 septembre 2005. Il est impératif d'y joindre une copie de l'article analysé. Les projets peuvent être déposés dans la boîte aux lettres de la chaire de Statistique Appliquée (accès 17-0-12) ou adressés à G.Saporta, Chaire de Statistique Appliquée, case 441, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris cedex 03. N'oubliez pas de mentionner votre nom, une adresse électronique ou un numéro de téléphone. Une présentation orale de la note de lecture pourra être demandée dans la semaine du 19 septembre.

Articles

I. Ph. CASIN - L'analyse discriminante de tableaux évolutifs, Revue de Statistique Appliquée, Vol. XLIII, N° 3, 1995 (attribué à Christine LEONARD-POYAU)

2. P. CAZES, A. CHOUAKRIA, E. DIDAY, Y. SCHEKTMAN - Extension de l'Analyse en Composantes Principales à des données de type intervalle, Revue de Statistique Appliquée. Vol. XLV, N° 3, 1997 (attribué à Désiré Seu)
3. F. FERRATY. - Estimation de transformations optimales en ACP curvilinéaire, Revue de Statistique Appliquée, Vol. XLV, N° 1, 1997 (attribué à G.Demoulin)
4. J. PAGES. Eléments de comparaison entre l'Analyse Factorielle Multiple et la méthode STATIS, Revue de Statistique Appliquée, Vol. XLIV, N° 4, 1996 (attribué à Amina Bouras)
5. M. TENENHAUS - L'approche PLS, Revue de Statistique Appliquée, Vol. XLVII, N° 2, 1999 (attribué à Fairouz Mohatta)
6. A.HYVARINEN, E.OJA – Independent component analysis: algorithms and applications, Neural networks, vol 13, 411-430, 2000, http://www.cis.hut.fi/aapo/papers/IJCNN99_tutorialweb/ (attribué à Patrick Morose)
7. B.SCHOLKOPF, A.SMOLA, K.R. MULLER- Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation, 10:1299-1319, 1998. Technical Report No. 44, 1996, Max Planck Institut für biologische Kybernetik, Tübingen. <http://www.kernel-machines.org/papers/nlpca.ps.gz> (attribué à Elie-Pierre Saint-Cyr)
8. J. PAGES, Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes, Revue de Statistique Appliquée, vol L, n°4, 5-37 (2002) (attribué à Boris Guannel)
9. X. BRY Une autre approche de l'analyse factorielle. L'analyse en composantes covariantes, Revue de Statistique Appliquée, vol XLIX, n°3, 5-38 (2001) (attribué à Hung-Feng Yu)
10. J.O. RAMSAY, Some statistical approaches to multidimensional scaling data, Journal of the Royal Statistical Society A, 145, 285-312 (1982). Avec la discussion. (attribué à H.Parthenay)
11. Abba M. KRIEGER, Paul E. GREEN - A Generalized Rand-Index Method for Consensus Clustering of Separate Partitions of the Same Data Base, J. of Classification 16:63-89 (1999) (attribué à M.Songsons)
12. A.D. GORDON, M. VICHI - Partitions of Partitions, J. of Classification 15:265-285 (1998) (attribué à M.Magassouba)
13. G. DIANA, C. TOMMASI – Cross validation methods in principal component analysis: a comparison., Statistical Methods and Applications, vol 11, n°1, 71-82 (2002) (attribué à Auguste Mpacko Priso)
14. Asa Ben-Hur, David Horn, Hava T. Siegelmann, Vladimir Vapnik, Support Vector Clustering, Journal of Machine learning Research, 2:125-137, 2001. <http://www.ai.mit.edu/projects/jmlr/papers/volume2/horn01a/rev1/abstract.html>

15. Règles d'association :

http://www.infres.enst.fr/~hebrail/publications/HI23_99_027.pdf (attribué à Michel Fernandes)

16. V. NZOBOUNSA, T. DHORNE : ECART : Une nouvelle méthode d'Analyse Canonique Généralisée, Revue de Statistique Appliquée, vol LI, n°4, 57-82 (2003) (attribué à Gael Cesbron)

17. Van Gesteel, and al. A support vector machine approach to credit scoring, <http://www.geocities.com/joaogarcia18/BANKFINVer4.pdf> (attribué à J.Tubiana)

18. Thaddeus Tarpey, Kimberly K. J. Kinader : Clustering Functional Data , Journal of Classification, 20, n°1 pp. 93 - 114 (2003) (attribué à Nicolas Simon)

19 A.Hardy : On the number of clusters, Computational Statistics and Data Analysis, Volume 23, Issue 1 Pages 83-96, (1996)

20. Hans H. Bock : Probabilistic models in cluster analysis, Computational Statistics and Data Analysis, Volume 23, Issue 1 Pages 5-28 (1996)

21. Roman Rosipal, Leonard J. Trejo: Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space , JMLR Special Issue on Kernel Methods, 2001 <http://www.jmlr.org/papers/v2/rosipal01a.html> (attribué à Akli Haddak)

22. Banasik, Crook : Does Reject Inference really improve the Performance of Application Scoring Models? <http://www.crc.man.ed.ac.uk/workingpapers/workingpaper02-3.pdf> (2002) (attribué à Jean-Paul Félix)

23 A.Buja, D.Swayne: Visualization Methodology for Multidimensional Scaling, Journal of Classification, Volume 19, Number 1, 2002, Pages: 7 – 43

24 Total least squares <ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/markovsky/reports/05-34.ps.gz>

Logiciels

1. Procédure MODECLUS de SAS (attribué à Rémi Saucié et Pierre-Yves Antier)

2. Procédure MDS de SAS (attribué à MARTINEAU Nicolas et FAYE-CHELLALI Sami)

3. Procédure VARCLUS de SAS (attribué à Laure Rodriguez et Vilay HOMPOMSILTHAM.)

4. Procédure PLS de SAS (attribué à Bruno Vivan et Sandrine Massé)

5. Procédures MI et MIANALYZE de SAS (attribué d'une part à Steve MambouNguie et à Sabine Suire et Thomas Rivron d'autre part)

6. Procédure GAM de SAS (attribué à M.El Ghebzouri)

7. Procédure FACTOR de SAS (attribué à Sebastien Nouet et El Abdellaoui Mustapha)

La documentation de SAS V9 est accessible à l'adresse:

<http://support.sas.com/onlinedoc/913/docMainpage.jsp>

8. Cartes de Kohonen. <http://samos.univ-paris1.fr/cours.html> et macros SAS

<http://samos.univ-paris1.fr/logiciel.html>, <http://eric.univ-lyon2.fr/~ricco/tanagra>

(Attribué à Messaoud CHEBEL et Mourad BENTOUMI)

9. Arbres et graphes de décision: SIPINA et TANAGRA <http://eric.univ-lyon2.fr/~ricco/sipina.html> <http://eric.univ-lyon2.fr/~ricco/tanagra> (attribué à S.Roesberg et E.Daupley)

10. Classes latentes : logiciel LatentGold version de démo gratuite téléchargeable à l'adresse : <http://www.statisticalinnovations.com/> (attribué à Christophe Vallet)