

Cours d'Analyse Multivariée Approfondie
STA 201
2010-2011

Projet

Au choix : un commentaire d'article ou une étude de logiciel.

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 ni 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 convient 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 (autres que les jeux d'essais fournis avec le logiciel).

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. Conseil : regardez les articles avant de choisir...

Remise du travail :

Au plus tard le 15 septembre 2011. Aucun report ne sera accepté. Dans le cas d'une note de lecture d'un article, il est **impératif** d'y joindre une copie de l'article analysé. Les rapports seront envoyés de préférence par courrier électronique sous forme d'un **fichier unique** au format PDF à gilbert.saporta@cnam.fr, ou par courrier postal à G.Saporta : Chaire de Statistique Appliquée, case 441, 292 rue Saint Martin, 75141 Paris cedex 03

Articles

La plupart des articles sont téléchargeables depuis les postes informatique du CNAM, sinon par accès à distance aux ressources électroniques : <http://bibliotheque.cnam.fr/ressources-electroniques/comment-se-connecter-a-distance/>

Pour la Revue de Statistique Appliquée, le site est <http://www.numdam.org/numdam-bin/feuilleter?j=RSA&sl=0>

1. Donald R. Jensen and Donald E. Ramirez: Anomalies in the Foundations of Ridge Regression, *International Statistical Review*, Volume 76, Issue 1, April 2008, pages 89–105 **suivi de** Prasenjit Kapat, Prem K. Goel : Anomalies in the Foundations of Ridge Regression: Some Clarifications, *International Statistical Review*, Volume 78, Issue 2, pages 209–215, August 2010 **et** Donald R. Jensen, Donald E. Ramirez : Anomalies in Ridge Regression: Rejoinder, *International Statistical Review*, Volume 78, Issue 2, pages 215–217, August 2010 **Exceptionnellement cet ensemble de 3 articles peut faire l'objet d'un rapport rédigé à deux.**
Attribué à Denis Guélen

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 Mohamed-Lamine SYLLA

3. F. FERRATY. - Estimation de transformations optimales en ACP curvilinéaire, *Revue de Statistique Appliquée*, Vol. XLV, N° 1, 1997 attribué à Serge AKAI

4. J. PAGÈS. Eléments de comparaison entre l'Analyse Factorielle Multiple et la méthode STATIS, *Revue de Statistique Appliquée*, Vol. XLIV, N° 4, 1996 attribué à Aurelia Petit

5. J.PAGÈS , Analyse factorielle de données mixtes, *Revue de Statistique Appliquée*, 52, n°4, 2004 attribué à Patrice Landel

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 Nicolas Hornn

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 Fabien Schaetzel

8. N. Louw and S.J. Steel: variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination *Computational Statistics & Data Analysis*, Volume 51, Issue 3, December 2006, Pages 2043-2055

9. Jean-François DURAND :La régression Partial Least-Squares boostée. La revue *MODULAD*, numéro 38, 2008. <http://www-rocq.inria.fr/axis/modulad/archives/numero-38/Durand-38/durand.pdf> attribué à Bruno Launay

10. P.BASTIEN, M.TENENHAUS, V.ESPOSITO-VINZI : PLS Generalized linear regression , Computational Statistics and Data Analysis, 48, 17-46, (2005) attribué à Dalice FERET ANVELET

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 Dominique Sorbier

12. A.D. GORDON, M. VICHI - Partitions of Partitions, J. of Classification 15:265-285 (1998) attribué à Anne LIU

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 Vivien El Jalti

14. A.BEN-HUR, D.HORN, H.T. SIEGELMANN, V. 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> attribué à Jean-Luc Vaillant

15. Sugnet Gardner and Niël J. le Roux, Extensions of Biplot Methodology to Discriminant Analysis, Journal of Classification, 22, n°1,pp 59 – 86, (2005)

16. Cathy W.S. Chen ,Jennifer S.K. Chan , Mike K.P. So and Kevin K.M. Lee: Classification in segmented regression problems, Computational Statistics & Data Analysis Volume 55, Issue 7, 1 July 2011, Pages 2276-2287

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

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

19. Eric J. Beh and Luigi D’Ambra :Some Interpretative Tools for Non-Symmetrical Correspondence Analysis, Journal of Classification, Volume 26, Number 1 / avril 2009, 55-76

20. Joost van Rosmalen, Patrick J. F. Groenen, Javier Trejos and William Castillo: Optimization Strategies for Two-Mode Partitioning, Journal of Classification, Volume 26, Number 2 / août 2009, 155-181

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>

22. Donatella Vicari and Maurizio Vichi :Structural Classification Analysis of Three-Way Dissimilarity Data, Journal of Classification, Volume 26, Number 2 / août 2009, 121-154

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

24. Mingan Yang and David B. Dunson : Bayesian Semiparametric Structural Equation Models with Latent Variables, Psychometrika, Volume 75, Number 4, 675-693 attribué à César Zinga

25. Michel van de Velden, Optimal Scaling of Paired Comparison Data , Journal of Classification, 21, n°1 pp89 – 109, (2004)

26. Jieping Ye, Characterization of a Family of Algorithms for Generalized Discriminant Analysis on Undersampled Problems , Journal of Machine Learning Research, 6:483--502, 2005. , <http://jmlr.csail.mit.edu/papers/volume6/ye05a/ye05a.pdf> attribué à Abdelbassit Khalifi

27. Ana M. Aguilera, Manuel Escabias and Mariano J. Valderrama , Using principal components for estimating logistic regression with high-dimensional multicollinear data, Computational Statistics & Data Analysis, Volume 50, Issue 8, Pages 1905-1924, (2006)

28. S. Bougeard, M. Hanafi and E.M. Qannari: Continuum redundancy–PLS regression: A simple continuum approach, Computational Statistics & Data Analysis, volume 52, n°7, (2008) pages 3686-3696 attribué à Mariam Farjallah

29. Raffaella Piccarreta : Classification trees for ordinal variables, Computational Statistics, Volume 23, Number 3 (2008) pages 407-427 attribué à Jacobus Kraakman

30. Joaquín Pacheco, Silvia Casado, Laura Núñez and Olga Gómez
Analysis of new variable selection methods for discriminant analysis • Computational Statistics & Data Analysis, Volume 51, n°3, Pages 1463-1478

31. António Pedro Duarte Silva and Paula Brito , Linear discriminant analysis for interval data, Computational Statistics, Volume 21, Number 2 / June, 2006 , 289-308

32. Govaert, G., Classification binaire et modèles. Revue de Statistique Appliquée, 38 no. 1 (1990), p. 67-81 attribué à Chaïness TALEB

33. Marcotorchino, F.; Michaud, P. Agrégation de similarités en classification automatique. Revue de Statistique Appliquée, 30 no. 2 (1982), p. 21-44 attribué à Gwenhael Billiotte

34. Celeux, G. Classification et modèles. Revue de Statistique Appliquée, 36 no. 4 (1988), p. 43-57 attribué à GAFA Kwami

35. M. Escabias, A.M. Aguilera, M.J. Valderrama :Functional PLS logit regression model. Computational Statistics & Data Analysis, Volume 51, Issue 10, 15 June 2007, Pages 4891-4902

36. Michael J. Brusco, Renu Singh and Douglas Steinley : Variable Neighborhood Search Heuristics for Selecting a Subset of Variables in Principal Component Analysis, Psychometrika, Volume 74, Number 4 / décembre 2009, 705-726

37. Heungsun Hwang: Regularized Generalized Structured Component Analysis, Psychometrika, Volume 74, Number 3, 517-530

38. Ke-Hai Yuan: Identifying Variables Responsible for Data not Missing at Random; Psychometrika, Volume 74, Number 2, 233-256 attribué à THACH Dara

39. Douglas Steinley and Michael J. Brusco: Selection of Variables in Cluster Analysis: An Empirical Comparison of Eight Procedures, Psychometrika, Volume 73, Number 1 / mars 2008, 125-144

40. Pier Alda Ferrari and Silvia Salini: Complementary Use of Rasch Models and Nonlinear Principal Components Analysis in the Assessment of the Opinion of Europeans About Utilities, Journal of Classification, DOI: 10.1007/s00357-011-9081-0 attribué à Ait Yahia Khalid

41. Rosaria Lombardo and Jacqueline J. Meulman: Multiple Correspondence Analysis via Polynomial Transformations of Ordered Categorical Variables, Journal of Classification Volume 27, Number 2, 191-210

42. Mark Ming-Tso Chiang and Boris Mirkin : Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads Journal of Classification, Volume 27, Number 1, 3-40

43. Kohei Adachi: Constrained principal component analysis of standardized data for biplots with unit-length variable vectors, Advances in Data Analysis and Classification Volume 5, Number 1, 23-36

44. Hye Won Suk and Heungsun Hwang : Regularized fuzzy clusterwise ridge regression, Advances in Data Analysis and Classification, Volume 4, Number 1, 35-51

45. Véronique Cariou, Stéphane Verdun, Emmanuelle Diaz, El Mostafa Qannari and Evelyne Vigneau: Comparison of three hypothesis testing approaches for the selection of the appropriate number of clusters of variables, :Advances in Data Analysis and Classification Volume 3, Number 3, 227-241 attribué à Geoffroy de Beaucorps

Logiciels

1. Procédure MODECLUS de SAS attribué à Nivoharilala RAZANAMPARANY et Yoland JOSEPH
2. Procédure ACECLUS de SAS attribué à Laetitia Spanu & Stéphanie Schultz
3. Procédure MDS de SAS attribué à Lionel Texier et Victor Chambaud
4. Procédure VARCLUS de SAS attribué à Paul Anh Kiet HOANG
5. Procédure PLS de SAS attribué à Yong ZHANG
6. Procédure GAM de SAS attribué à Idriss AMRANI et Karine Darjinoff
7. Procédure FACTOR de SAS attribué à Nicolas Foubert de Pallières et Abelkrim Rebadj.
8. Procédure LOESS de SAS attribué à ROMAND-PIQUANT et JEGOU Christian
9. Procédure KDE de SAS attribué à Alex SY et Martine Lecamus
10. Procédures PHREG et TPHREG de SAS attribué à lebamba-moussa et ousmane rhesus
11. Procédures MI et MIANALYZE de SAS attribué à Julie Leblanc
12. Procédure ROBUSTREG de SAS attribué à Goumar Ali et Bouddour Hicham
13. Procédure QUANTREG de SAS attribué à MPAN SIMPLICE LANDRY
14. Procédure PRINQUAL de SAS attribué à Anne-Laure GENIES LIAGRE et Fabien NANTEUIL
15. Procédure TRANSREG de SAS attribué à JEAN-ELIE Marie-Hélène
16. 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 BENMALEK Mohand
17. Classes latentes : logiciel LatentGold version de démo gratuite téléchargeable à l'adresse : <http://www.statisticalinnovations.com/> attribué à Fabien Piocelle
18. Classes latentes : PROC LCA développée à Penn State University : <https://methodology.psu.edu/downloads/proclcalta> attribué à Hamza Rabhi et Eric Zahui