zbMATH Open
THE FIRST RESOURCE FOR MATHEMATICS

Dean, Nema; **Nugent, Rebecca**

Clustering student skill set profiles in a unit hypercube using mixtures of multivariate betas.

(English) Zbl 1416.62334
Adv. Data Anal. Classif., ADAC 7, No. 3, 339-357 (2013).

Summary: This paper presents a finite mixture of multivariate betas as a new model-based clustering method tailored to applications where the feature space is constrained to the unit hypercube. The mixture component densities are taken to be conditionally independent, univariate unimodal beta densities (from the subclass of reparameterized beta densities given by *L. Bagnato* and *A. Punzo* [Comput. Stat. 28, No. 4, 1571–1597 (2013; Zbl 1306.65024)]. The EM algorithm used to fit this mixture is discussed in detail, and results from both this beta mixture model and the more standard Gaussian model-based clustering are presented for simulated skill mastery data from a common cognitive diagnosis model and for real data from the Assistment System online mathematics tutor [*M. Feng* et al., "Addressing the assessment challenge with an online system that tutors as it assesses", User Model. User-Adapted Interact. 19, No. 3, 243–266 (2009; doi:10.1007/s11257-009-9063-7)]. The multivariate beta mixture appears to outperform the standard Gaussian model-based clustering approach, as would be expected on the constrained space. Fewer components are selected (by BIC-ICL) in the beta mixture than in the Gaussian mixture, and the resulting clusters seem more reasonable and interpretable.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)

Cited in **3** Documents

Keywords:

mixture model clustering; multivariate beta densities; skill set profiles; unit hypercube

Software:

R

Full Text: DOI

References:

[1] Ayers E, Nugent R, Dean N (2008) Skill set profile clustering based on student capability vectors computed from online tutoring data. In: Baker R, Barnes T, Beck JE (eds) Proceedings of the 1st international conference on educational data mining. Montreal, Canada, pp 210–217

[2] Ayers E, Nugent R, Dean N (2009) A comparison of student skill knowledge estimates. In: Barnes T, Desmarais M, Romero C, Ventura S (eds) Proceedings of the 2nd international conference on educational data mining. Cordoba, Spain, pp 1–10

[3] Bagnato L, Punzo A (2013) Finite mixtures of unimodal beta and gamma densities and the \$\$k\$\$ k -bumps algorithm. Computational Statistics 28(4): doi: 10.1007/s00180-012-367-4 · Zbl 1306.65024

[4] Barnes TM (2005) The Q-matrix method: mining student response data for knowledge. In: Beck JE (ed) Educational data mining: papers from the 2005 AAAI workshop. American Association for Artificial Intelligence, Menlo Park, California, Technical, Report WS-05-02, pp 39–46

[5] Baudry JP, Raftery AE, Celeux G, Lo K, Gottardo R (2010) Combining mixture components for clustering. J Comput Graph Stat 19(2):332–353· doi:10.1198/jcgs.2010.08111

[6] Dean N, Nugent R (2011) Comparing different clustering models on the unit hypercube. In: Proceedings of the 58th world statistics congress. International Statistical Institute, Dublin

[7] Dean N, Nugent R (2013) Mixture model component trees: Visualizing the hierarchical structure of complex groups. Tech. rep., University of Glasgow (in preparation) · Zbl 1416.62334

[8] Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc Ser B Methodol 39(1):1–38 with discussion · Zbl 0364.62022

[9] DiBello L, Roussos L, Stout W (2007) Review of cognitively diagnostic assessment and a summary of psychometric models. In: Rao CR, Sinharay S (eds) Handbook of Statistics, 26. Elsevier, Amsterdam, pp 979–1030 · Zbl 1255.91382

[10] Feng M, Heffernan N, Koedinger K (2009) Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. J User Model User Adapt Inter 19(3):243–266 · Zbl 05672262 · doi:10.1007/s11257-009-9063-7

© 2022 FIZ Karlsruhe GmbH

[11] Fraley C, Raftery AE (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput J 41:578–588 · Zbl 0920.68038 · doi:10.1093/comjnl/41.8.578

[12] Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 97(458):611–612 · Zbl 1073.62545 · doi:10.1198/016214502760047131

[13] Fraley C, Raftey AE (2007) MCLUST version 3 for R: normal mixture modeling and model-based clustering. Tech. Rep. 504, Department of Statistics, University of Washington, Washington

[14] Fraley C, Raftey AE, Murphy TB, Scrucca L (2012) mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation. Tech. Rep. 597, Department of Statistics, University of Washington, Washington

[15] Hennig C (2010a) Methods for merging Gaussian mixture components. Adv Data Anal Class 4(1):3–34 · Zbl 1306.62141 · doi:10.1007/s11634-010-0058-3

[16] Hennig C (2010b) Ridgeline plot and clusterwise stability as tools for merging Gaussian mixture components. In: Locarek-Junge H, Weihs C (eds) Classification as a tool for research. Springer, Berlin, pp 109–116

[17] Henson J, Templin R, Douglas J (2007) Using efficient model based sum-scores for conducting skill diagnoses. J Edu Measur 44(4):361–376· doi:10.1111/j.1745-3984.2007.00044.x

[18] Hubert L, Arabie P (1985) Comparing partitions. J Class 2(1):193–218 · Zbl 0587.62128 · doi:10.1007/BF01908075

[19] Ji Y, Wu C, Liu P, Wang J, Coombes KR (2005) Applications of beta-mixture models in bioinformatics. Bioinformatics 21(9):2118–2122· doi:10.1093/bioinformatics/bti318

[20] Junker BW, Sijtsma K (2001) Cognitive assessment models with few assumptions and connections with nonparametric item response theory. Appl Psych Meas 25(3):258–272· doi:10.1177/01466210122032064

[21] Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795 · Zbl 0846.62028 · doi:10.1080/01621459.1995.10476572

[22] Lazarsfeld PF, Henry PW (1968) Latent structure analysis. Houghton Mifflin, Boston

[23] Lindsay BG (1995) Mixture models: theory, geometry, and applications. In: NSF-CBMS regional conference series in probability and statistics, vol. 5, Institute of Mathematical Statistics, Hayward

[24] Maugis C, Celeux G, Martin-Magniette ML (2009) Variable selection in model-based clustering: a general variable role modeling. Comput Stat Data Anal 53(11):3872–3882 · Zbl 1453.62154 · doi:10.1016/j.csda.2009.04.013

[25] McLachlan G, Peel D (1999) The EMMIX algorithm for the fitting of normal and t-components. J Stat Softw 4(2):1–14

[26] McLachlan GJ, Peel D (2000) Finite mixture models. Wiley, New York · Zbl 0963.62061

[27] Nugent R, Ayers E, Dean N (2009) Conditional subspace clustering of skill mastery: identifying skills that separate students. In: Barnes T, Desmarais M, Romero C, Ventura S (eds) Proceedings of the 2nd international conference on educational data mining. Cordoba, Spain, pp 101–110

[28] R Core Team (2012) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna. http://www.R-project.org/ , ISBN 3-900051-07-0

[29] Raftery AE, Dean N (2006) Variable selection for model-based clustering. J Am Stat Assoc 101(473): 168–178 · Zbl 1118.62339 · doi:10.1198/016214506000000113

[30] Rupp AA, Templin J, Henson RA (2010) Diagnostic measurement: theory, methods, and applications. Guilford Press, New York

[31] Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464 · Zbl 0379.62005 · doi:10.1214/aos/1176344136

[32] Sokal RR, Rohlf JF (1981) Biometry: the principles and practice of statistics in biological research, 2nd edn. W. H Freemand and Company, San Francisco · Zbl 0554.62094

[33] Torre JDL (2009) DINA model and parameter estimation: a didactic. J Edu Behav Stat 34(1):115–130· doi:10.3102/1076998607309474

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.

© 2022 FIZ Karlsruhe GmbH