zbMATH Open
THE FIRST RESOURCE FOR MATHEMATICS

Ahfock, Daniel; McLachlan, Geoffrey J.
Harmless label noise and informative soft-labels in supervised classification. (English)
Zbl 07422751
Comput. Stat. Data Anal. 161, Article ID 107253, 12 p. (2021)

Summary: Manual labelling of training examples is common practice in supervised learning. When the labelling task is of non-trivial difficulty, the supplied labels may not be equal to the ground-truth labels, and label noise is introduced into the training dataset. If the manual annotation is carried out by multiple experts, the same training example can be given different class assignments by different experts, which is indicative of label noise. In the framework of model-based classification, a simple, but key observation is that when the manual labels are sampled using the posterior probabilities of class membership, the noisy labels are as valuable as the ground-truth labels in terms of statistical information. A relaxation of this process is a random effects model for imperfect labelling by a group that uses approximate posterior probabilities of class membership. The relative efficiency of logistic regression using the noisy labels compared to logistic regression using the ground-truth labels can then be derived. The main finding is that logistic regression can be robust to label noise when label noise and classification difficulty are positively correlated. In particular, when classification difficulty is the only source of label errors, multiple sets of noisy labels can supply more information for the estimation of a classification rule compared to the single set of ground-truth labels.

MSC:

62-XX Statistics

Keywords:

label noise; multiple labels; model-based classification; supervised classification

Software:

MBCbook; UCI-ml

Full Text: DOI arXiv

References:

[1] Bi, Y.; Jeske, D. R., The efficiency of logistic regression compared to normal discriminant analysis under class-conditional classification noise, J. Multivar. Anal., 101, 1622-1637 (2010) · Zbl 1189.62104

[2] Blanchard, G.; Flaska, M.; Handy, G.; Pozzi, S.; Scott, C., Classification with asymmetric label noise: consistency and maximal denoising, Electron. J. Stat., 10, 2780-2824 (2016) · Zbl 1347.62106

[3] Bouveyron, C.; Celeux, G.; Murphy, T.; Raftery, A., Model-Based Clustering and Classification for Data Science: With Applications in R (2019), Cambridge University Press · Zbl 1436.62006

[4] Bouveyron, C.; Girard, S., Robust supervised classification with mixture models: learning from data with uncertain labels, Pattern Recognit., 42, 2649-2658 (2009) · Zbl 1175.68313

[5] Cannings, T. I.; Fan, Y.; Samworth, R. J., Classification with imperfect training labels, Biometrika, 107, 311-330 (2020) · Zbl 1441.62165

[6] Cappozzo, A.; Greselin, F.; Murphy, T. B., A robust approach to model-based classification based on trimming and constraints, Adv. Data Anal. Classif., 14, 327-354 (2019) · Zbl 1474.62215

[7] Cheng, R. C.H.; Liu, W. B., The consistency of estimators in finite mixture models, Scand. J. Stat., 28, 603-616 (2001) · Zbl 1010.62023

[8] Clemmensen, L.; Hastie, T.; Witten, D.; Ersbøll, B., Sparse discriminant analysis, Technometrics, 53, 406-413 (2011)

[9] Dawid, A. P.; Skene, A. M., Maximum likelihood estimation of observer error-rates using the EM algorithm, J. R. Stat. Soc., Ser. C, Appl. Stat., 28, 20-28 (1979)

[10] Donmez, P.; Carbonell, J.; Schneider, J., A probabilistic framework to learn from multiple annotators with time-varying accuracy, (Proceedings of the 2010 SIAM International Conference on Data Mining (2010), SIAM), 826-837

[11] Dua, D.; Graff, C., UCI Machine Learning Repository (2017), University of California, School of Information and Computer Sciences: University of California, School of Information and Computer Sciences Irvine

[12] Efron, B., The efficiency of logistic regression compared to normal discriminant analysis, J. Am. Stat. Assoc., 70, 892-898 (1975) · Zbl 0319.62039

[13] Efron, B.; Tibshirani, R., Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Stat. Sci., 1, 54-75 (1986) · Zbl 0587.62082

[14] Fahrmeir, L., Maximum likelihood estimation in misspecified generalized linear models, Statistics, 21, 487-502 (1990) · Zbl 0714.62066

[15] Fahrmeir, L.; Kaufmann, H., Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, Ann. Stat., 13, 342-368 (1985) · Zbl 0594.62058

[16] Fraley, C.; Raftery, A. E., Model-based clustering, discriminant analysis, and density estimation, J. Am. Stat. Assoc., 97, 611-631 (2002) · Zbl 1073.62545

[17] Frénay, B.; Verleysen, M., Classification in the presence of label noise: a survey, IEEE Trans. Neural Netw. Learn. Syst., 25, 845-869 (2014)

[18] Hovy, D.; Berg-Kirkpatrick, T.; Vaswani, A.; Hovy, E., Learning whom to trust with MACE, (Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (2013), Association for Computational Linguistics), 1120-1130

[19] Ipeirotis, P. G.; Provost, F.; Wang, J., Quality management on Amazon Mechanical Turk, (Proceedings of the ACM SIGKDD Workshop on Human Computation (2010)), 64-67

[20] Jin, R.; Ghahramani, Z., Learning with multiple labels, (Advances in Neural Information Processing Systems (2003)), 921-928

[21] Johnson, N. L.; Kotz, S.; Balakrishnan, N., Discrete Multivariate Distributions (1997), Wiley: Wiley New York · Zbl 0868.62048

[22] McLachlan, G. J., Asymptotic results for discriminant analysis when the initial samples are misclassified, Technometrics, 14, 415-422 (1972) · Zbl 0244.62046

[23] McLachlan, G. J.; Peel, D., Finite Mixture Models (2000), Wiley · Zbl 0963.62061

[24] Mesejo, P.; Pizarro, D.; Abergel, A.; Rouquette, O.; Beorchia, S.; Poincloux, L.; Bartoli, A., Computer-aided classification of gastrointestinal lesions in regular colonoscopy, IEEE Trans. Med. Imaging, 35, 2051-2063 (2016)

[25] Michalek, J. E.; Tripathi, R. C., The effect of errors in diagnosis and measurement on the estimation of the probability of an event, J. Am. Stat. Assoc., 75, 713-721 (1980) · Zbl 0446.62062

[26] Natarajan, N.; Dhillon, I. S.; Ravikumar, P. K.; Tewari, A., Learning with noisy labels, (Advances in Neural Information Processing Systems (2013)), 1196-1204

[27] Ng, A. Y.; Jordan, M. I., On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes, (Advances in Neural Information Processing Systems (2002)), 841-848

[28] Quost, B.; Denœux, T.; Li, S., Parametric classification with soft labels using the evidential EM algorithm: linear discriminant analysis versus logistic regression, Adv. Data Anal. Classif., 11, 659-690 (2017) · Zbl 1414.62265

[29] Raykar, V. C.; Yu, S., Eliminating spammers and ranking annotators for crowdsourced labeling tasks, J. Mach. Learn. Res., 13, 491-518 (2012) · Zbl 1283.68300

[30] Raykar, V. C.; Yu, S.; Zhao, L. H.; Jerebko, A.; Florin, C.; Valadez, G. H.; Bogoni, L.; Moy, L., Supervised learning from multiple experts: whom to trust when everyone lies a bit, (Proceedings of the 26th Annual International Conference on Machine Learning (2009)), 889-896

[31] Rebbapragada, U.; Brodley, C. E., Class noise mitigation through instance weighting, (European Conference on Machine Learning (2007), Springer), 708-715

[32] Sheng, V. S.; Provost, F.; Ipeirotis, P. G., Get another label? Improving data quality and data mining using multiple, noisy labelers, (Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008)), 614-622

[33] Smyth, P.; Fayyad, U. M.; Burl, M. C.; Perona, P.; Baldi, P., Inferring ground truth from subjective labelling of Venus images, (Advances in Neural Information Processing Systems (1995)), 1085-1092

[34] Song, H.; Dai, R.; Raskutti, G.; Barber, R. F., Convex and non-convex approaches for statistical inference with class-conditional noisy labels, J. Mach. Learn. Res., 21, 1-58 (2020) · Zbl 07306869

[35] Vranckx, I.; Raymaekers, J.; Ketelaere, B. D.; Rousseeuw, P. J.; Hubert, M., Real-time discriminant analysis in the presence of label and measurement noise, Chemom. Intell. Lab. Syst., 208, Article 104197 pp. (2021)

[36] Welinder, P.; Branson, S.; Perona, P.; Belongie, S. J., The multidimensional wisdom of crowds, (Advances in Neural Information Processing Systems (2010)), 2424-2432

[37] Xu, L.; Crammer, K.; Schuurmans, D., Robust support vector machine training via convex outlier ablation, (Proceedings of the 21st National Conference on Artificial Intelligence AAAI'06 (2006), AAAI Press), 536-542

[38] Yan, Y.; Rosales, R.; Fung, G.; Schmidt, M.; Hermosillo, G.; Bogoni, L.; Moy, L.; Dy, J., Modeling annotator expertise: learning when everybody knows a bit of something, (Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (2010)), 932-939

[39] Zhang, P.; Cao, W.; Obradovic, Z., Learning by aggregating experts and filtering novices: a solution to crowdsourcing problems in bioinformatics, BMC Bioinform., 14, 1-8 (2013)