The false discovery rate for statistical pattern recognition

Clayton Scott, *University of Michigan*
Gowtham Bellala, *University of Michigan*
Rebecca Willett, *Duke University*

Abstract

The false discovery rate (FDR) and false nondiscovery rate (FNDR) have received considerable attention in the literature on multiple testing. These performance measures are also appropriate for classification, and in this work we develop generalization error analyses for FDR and FNDR when learning a classifier from labeled training data. Unlike more conventional classification performance measures, the empirical FDR and FNDR are not binomial random variables but rather a ratio of binomials, which introduces challenges not present in conventional formulations of the classification problem. We develop distribution-free uniform deviation bounds and apply these to obtain finite sample bounds and strong universal consistency. We also present a simulation study demonstrating the merits of variance-based bounds, which we also develop. In the context of multiple testing with FDR/FNDR, our framework may be viewed as a way to leverage training data to achieve distribution free, asymptotically optimal inference under the random effects model.

AMS 2000 subject classifications: Primary 62H30; secondary 68T05.

Keywords: Statistical learning theory, generalization error, false discovery rate.

Full Text: PDF

Scott, Clayton, Bellala, Gowtham, Willett, Rebecca, The false discovery rate for statistical pattern recognition, Electronic Journal of Statistics, 3, (2009), 651-677 (electronic). DOI: 10.1214/09-EJS363.

References

[1] Agarwal, S., Graepel, T., Herbrich, R., Har-Peled, S., and Roth, D. (2005). Generalization bounds for the area under the ROC curve. J. Machine Learning Research 6, 393–425. MR2249826

[2] Altman, D. G. and Bland, J. M. (1994). Diagnostic tests 2: predictive values. Brit. Med. J. 309, 102.

[3] Arlot, S., Blanchard, G., and Roquain, E. (2007). Resampling-based confidence regions and multiple tests for a correlated random vector. In Learning Theory: 20th Annual Conference on Learning Theory, COLT 2007, N. H. B. C. Gentile, Ed. Springer-Verlag, Heidelberg, 127–141. MR2397583

[4] Bach, F. R., Heckerman, D., and Horvitz, E. (2006). Considering cost asymmetry in learning classifiers. J. Machine Learning Research, 1713–1741. MR2274422

[5] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc B 57, 1, 289–300. MR1325392

[6] Bernstein, S. (1946). The Theory of Probabilities. Gastehizdat Publishing House, Moscow.

[7] Beygelzimer, A., Langford, J., and Ravikumar, P. (2009). Error-correcting tournaments. preprint.

[8] Blanchard, G. and Fleuret, F. (2007). Occam's hammer. In Learning Theory: 20th Annual Conference on Learning Theory, COLT 2007, N. H. B. C. Gentile, Ed. Springer-Verlag, Heidelberg, 112–126. MR2397582

[9] Bousquet, O., Boucheron, S., and Lugosi, G. (2004). Introduction to statistical learning theory. In Advanced Lectures in Machine Learning, O. Bousquet, U. Luxburg, and G. Rätsch, Eds. Springer, 169–207.

[10] Cannon, A., Howse, J., Hush, D., and Scovel, C. (2002). Learning with the Neyman-Pearson and min-max criteria. Tech. Rep. LA-UR 02-2951, Los Alamos National Laboratory.

[11] Chi, Z. and Tan., Z. (2008). Positive false discovery proportions: intrinsic bounds and adaptive control. Statistica Sinica 18, 3, 837–860. MR2440397

[12] Clémenccon, S., Lugosi, G., and Vayatis, N. (2008). Ranking and empirical risk minimization of U-statistics. Ann. Stat. 36, 2, 844–874. MR2396817

[13] Clémencon, S. and Vayatis, N. (2009a). Nonparametric estimation of the precision-recall curve. In to appear, Proc. 26th International Machine Learning Conference (ICML).

[14] Clémencon, S. and Vayatis, N. (2009b). Overlaying classifiers: a practical approach for optimal ranking. In Advances in Neural Information Processing Systems 21, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. 313–320.

[15] Devroye, L., Györfi, L., and Lugosi, G. (1996). A Probabilistic Theory of Pattern Recognition. Springer, New York. MR1383093

[16] Devroye, L. and Lugosi, G. (2001). Combinatorial Methods in Density Estimation. Springer, New York. MR1843146

[17] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. Ann. Stat. 32, 3, 962–994. MR2065195

[18] Durrett, R. (1991). Probability: Theory and Examples. Wadsworth & Brooks/Cole, Pacific Grove, CA. MR1068527

[19] Efron, B., Tibshirani, R., Storey, J., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. Journal of the American Statistical Association 96, 1151–1160. MR1946571

[20] El-Yaniv, R. and Pechyony, D. (2007). Transductive Rademacher complexity and its applications. In Proc. 20th Annual Conference on Learning Theory, COLT 2007, N. Bshouty and C. Gentile, Eds. Springer-Verlag, Heidelberg, 157–171. MR2397585

[21] Elkan, C. (2001). The foundations of cost-sensitive learning. In Proceedings of the 17th International Joint Conference on Artificial Intelligence. Seattle, Washington, USA, 973–978.

[22] Genovese, C. R., Lazar, N. A., and Nichols, T. E. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.

[23] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13–30. MR0144363

[24] Liu, B., Lee, W. S., Yu, P. S., and Li, X. (2002). Partially supervised classification of text documents. In Proc. 19th Int. Conf. Machine Learning (ICML). Sydney, Australia, 387–394.

[25] Mansour, Y. and McAllester, D. (2000). Generalization bounds for decision trees. In Proceedings of the Thirteenth Annual Conference on Computational Learning Theory,

N. Cesa-Bianchi and S. Goldman, Eds. Palo Alto, CA, 69–74.

[26] McAllester, D. (1999). Some PAC-Bayesian theorems. Machine Learning 37, 3, 355–363.

[27] Scott, C. (2007). Performance measures for Neyman-Pearson classification. IEEE Trans. Inform. Theory 53, 8, 2852–2863. MR2400500

[28] Scott, C. and Blanchard, G. (2009). Novelty detection: Unlabeled data definitely help. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS) 2009 (April 16-18), D. van Dyk and M. Welling, Eds. JMLR: W&CP 5, Clearwater Beach, Florida, 464–471.

[29] Scott, C. and Nowak, R. (2005). A Neyman-Pearson approach to statistical learning. IEEE Trans. Inform. Theory 51, 8, 3806–3819. MR2239000

[30] Scott, C. and Nowak, R. (2006). Learning minimum volume sets. J. Machine Learning Res. 7, 665–704. MR2274383

[31] Scott, C. D., Bellala, G., and Willett, R. (2007). Generalization error analysis for FDR controlled classification. In Proc. IEEE Workshop on Statistical Signal Processing. Madison, WI.

[32] Soric, B. (1989). Statistical discoveries and effect-size estimation. J. Amer. Statist. Assoc. 84, 608–610.

[33] Storey, J. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B 64, 479–498. MR1924302

[34] Storey, J. (2003). The positive false discovery rate: A Bayesian interpretation of the q-value. Annals of Statistics 31:6, 2013–2035. MR2036398

[35] Storey, J. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. Journal of the Royal Statistical Society, Series B 69, 347–368. MR2323757

[36] Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. Ann. Stat. 32, 1, 135–166. MR2051002

[37] Usunier, N., Amini, M., and Gallinari, P. (2005). A data-dependent generalisation error bound for the AUC. In Proc. ICML 2005 Workshop on ROC Analysis in Machine Learning.

[38] van Rijsbergen, C. J. (1979). Information Retrieval, 2nd ed. Butterworths, London.

Electronic Journal of Statistics. ISSN: 1935-7524