© 2013 FIZ Karlsruhe

io-port 06058576

Yıldız, Olcay Taner; Aslan, Özlem; Alpaydın, Ethem

Multivariate statistical tests for comparing classification algorithms.

Coello Coello, Carlos A. (ed.), Learning and intelligent optimization. 5th international conference, LION 5, Rome, Italy, January 17–21, 2011. Selected papers. Berlin: Springer (ISBN 978-3-642-25565-6/pbk). Lecture Notes in Computer Science 6683, 1-15 (2011).

Summary: The misclassification error which is usually used in tests to compare classification algorithms, does not make a distinction between the sources of error, namely, false positives and false negatives. Instead of summing these in a single number, we propose to collect multivariate statistics and use multivariate tests on them. Information retrieval uses the measures of precision and recall, and signal detection uses true positive rate (tpr) and false positive rate (fpr) and a multivariate test can also use such two values instead of combining them in a single value, such as error or average precision. For example, we can have bivariate tests for (precision, recall) or (tpr, fpr). We propose to use the pairwise test based on Hotelling's multivariate T^2 test to compare two algorithms or multivariate analysis of variance (MANOVA) to compare $L > 2$ algorithms. In our experiments, we show that the multivariate tests have higher power than the univariate error test, that is, they can detect differences that the error test cannot, and we also discuss how the decisions made by different multivariate tests differ, to be able to point out where to use which. We also show how multivariate or univariate pairwise tests can be used as post-hoc tests after MANOVA to find cliques of algorithms, or order them along separate dimensions.

doi:10.1007/978-3-642-25566-3_1