

EGU2020-19751

<https://doi.org/10.5194/egusphere-egu2020-19751>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## A-posteriori Analyses of Pattern Recognition Results

**Horst Langer**, Susanna Falsaperla, and Conny Hammer

Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Catania, Italy (horst.langer@ct.ingv.it)

Data-driven approaches applied to large and complex data sets are intriguing, however the results must be revised with a critical attitude. For example, a diagnostic tool may provide hints for a serious disease, or for anomalous conditions potentially indicating an impending natural risk. The demand of a high score of identified anomalies – true positives – comes together with the request of a low percentage of false positives. Indeed, a high rate of false positives can ruin the diagnostics. Receiver Operation Curves (ROC) allows us to find a reasonable compromise between the need of accuracy of the diagnostics and robustness with respect to false alerts.

In multiclass problems success is commonly measured as the score for which calculated and target classification of patterns matches at best. A high score does not automatically mean that a method is truly effective. Its value becomes questionable, when a random guess leads to a high score as well. The so called “Kappa Statistics” is an elegant way to assess the quality of a classification scheme. We present some case studies demonstrating how such a-posteriori analysis helps corroborate the results.

Sometimes an approach does not lead to the desired success. In these cases, a sound a-posteriori analysis of the reasons for the failure often provides interesting insights into the problem. Those problems may reside in an inappropriate definition of the targets, inadequate features, etc. Often the problems can be fixed just by adjusting some choices. Finally, a change of strategy may be necessary in order to achieve a more satisfying result. In the applications presented here, we highlight the pitfalls arising in particular from ill-defined targets and unsuitable feature selections.

The validation of unsupervised learning is still a matter of debate. Some formal criteria (e. g. Davies Bouldin Index, Silhouette Index or other) are available for centroid-based clustering where a unique metric valid for all clusters can be defined. Difficulties arise when metrics are defined individually for each single cluster (for instance, Gaussian Model clusters, adaptive criteria) as well as using schemes where centroids are essentially meaningless. This is the case in density based clustering. In all these cases, users are better off when asking themselves whether a clustering is meaningful for the problem in physical terms. In our presentation we discuss the problem of choosing a suitable number of clusters in cases in which formal criteria are not applicable. We demonstrate how the identification of groups of patterns helps the identification of elements which have a clear physical meaning, even when strict rules for assessing the clustering are not available.

