A-posteriori Analyses of Pattern Recognition Results

EGU General Assembly 2020

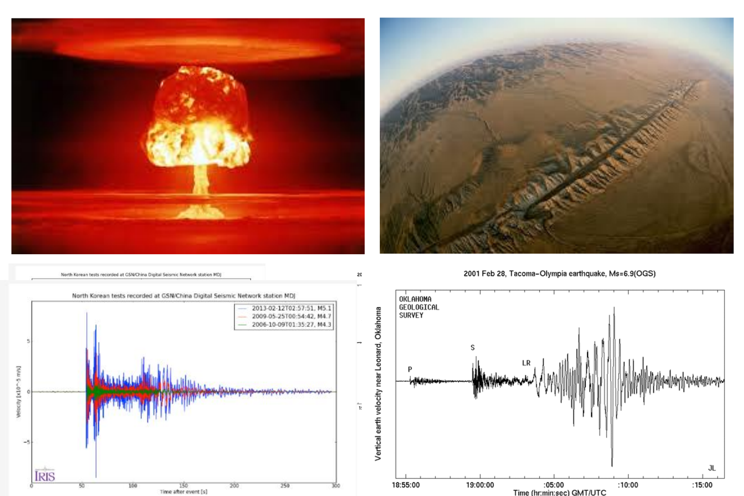
EGU2020-19751

Horst Langer[1], Susanna Falsaperla[1], Conny Hammer[2]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Catania, Osservatorio Etneo, Italy

[2]Swiss Seismological Service, Eidgenössische Technische Hochschule (ETH) Zurich, Switzerland

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

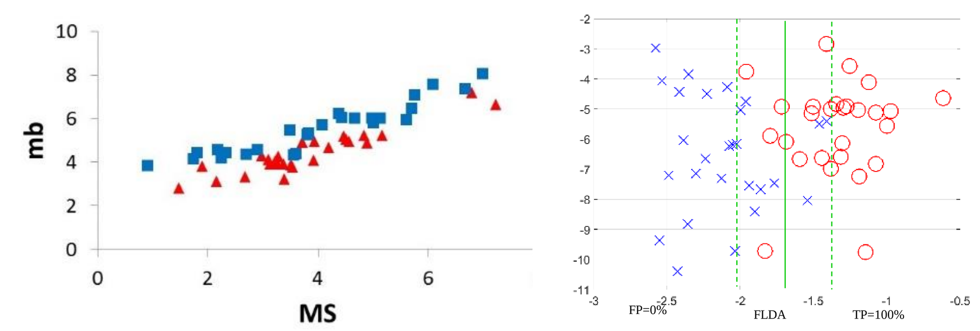


F1

In Geophysics we study phenomena, such as earthquakes, and nuclear tests, i.e., **Objects** (F1). Observables (seismograms) form **Patterns**. Pattern Recognition relates patterns to objects, e. g., seismograms to earthquakes or nuclear tests. Direct use of patterns is often not effective. Extracting **Features** reduces the amount of data.

Here we use body wave magnitudes – mb - and surface wave magnitudes - MS. Using Fisher's Discrimination we establish discrimination thresholds learned from examples (F2).

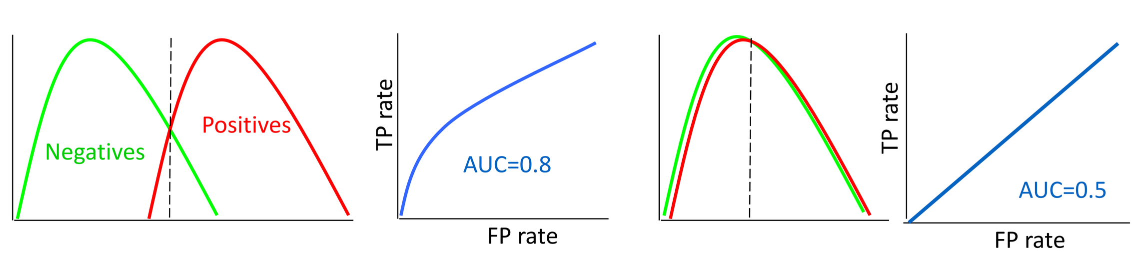
F2



Discovering all NUKEs ("True positive rate TP" = 100%) with some false detections of NUKE ("false positives" FP). Avoiding FP means that some NUKE is not detected.

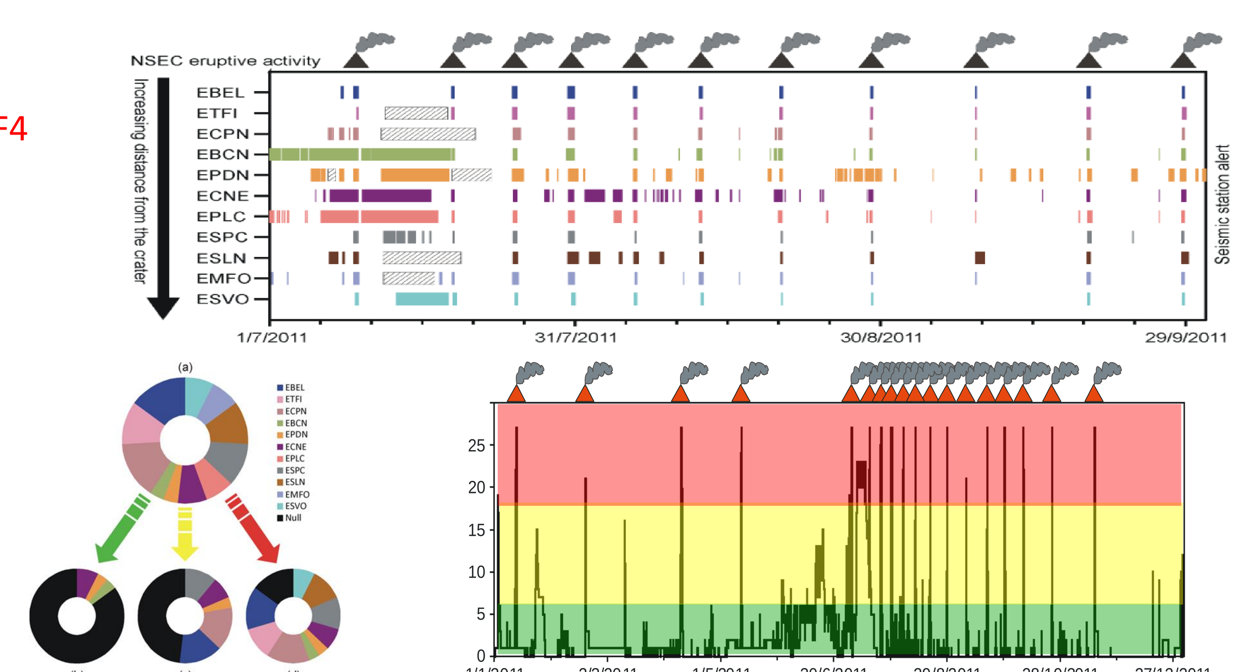
The threshold affects the number of true positives (TP) as well as the false positives (FP). In **Receiver Operation Curves** (ROC) we plot TP vs FP rates for varying thresholds (F3). The **Area Under the ROC** (AU) is a general parameter for the quality of discrimination (1 for full distinction, 0.5 for no distinction).

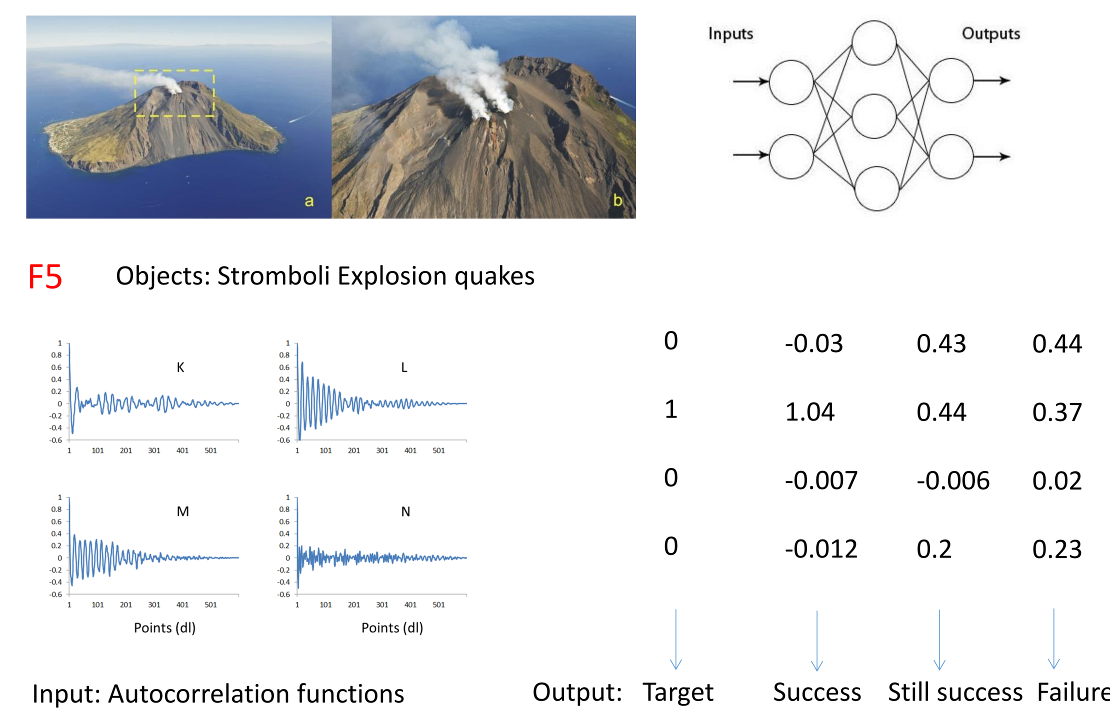
F3



The reasoning outlined above also applies to the decision whether to take action. In **Volcano Monitoring** on Mt Etna (F4) we exploit unsupervised pattern recognition to identify the unrest of the volcano considering seismic stations. We adopt a voting scheme based on the number of stations signaling a criticality. Then, we calculate ROC and AU. The definition of the threshold to be used remains a task involving both researchers as well as end users, who have to decide how many FP are tolerable.

F4

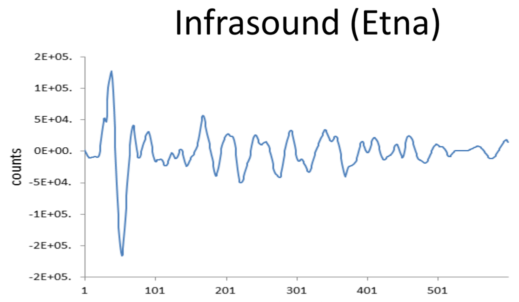




F5

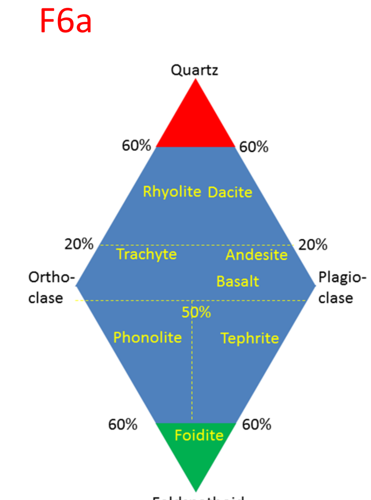
In multiclass problems the classifier learns to assign patterns to class A, B, C,…N. The success is assessed from a "confusion" matrix, reporting the calculated output of the classifier with respect to the target class. The success rate corresponds to the sum of matching classifications (diagonal elements in the confusion matrix).

Infrasound (Etna)



True | Predicted	Class 1	Class 2	Class 3
Class 1	476	9	6
Class 2	9	15	8
Class 3	8	33	46

F6a

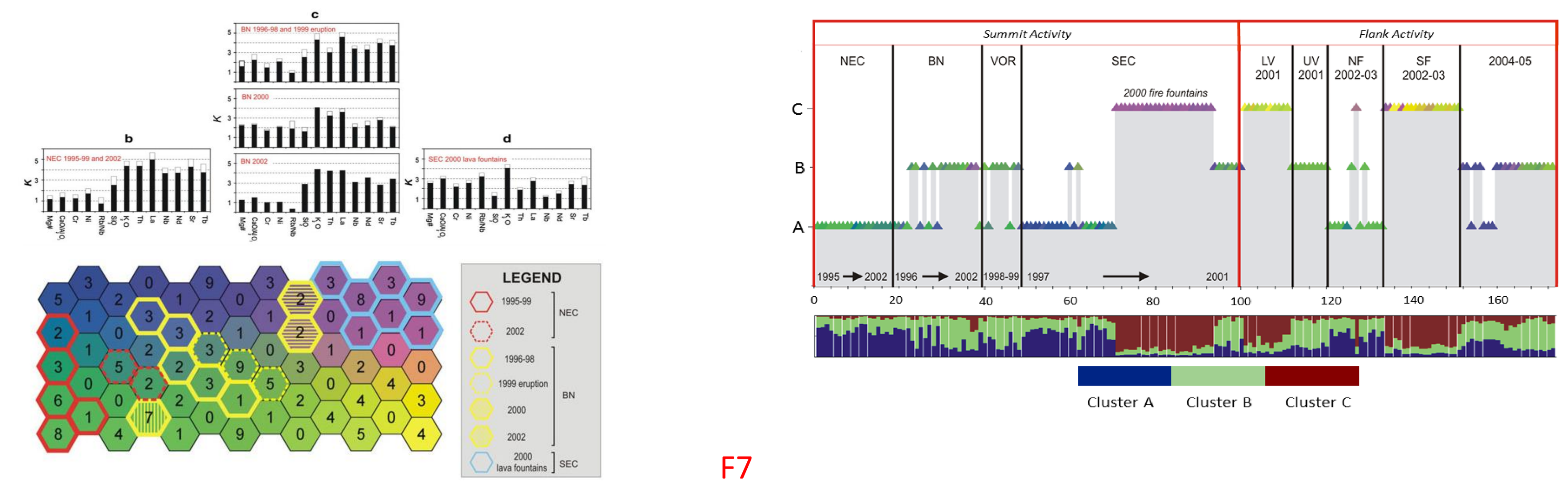


SiO_2	TiO_2	Al_2O_3	CaO	MgO	MnO	K_2O	Na_2O	P_2O_5	Rockname
6.64E+01	4.75E-01	1.44E+01	1.45E+00	3.00E-01	1.1E-01	5.80E+00	3.55E+00	7.00E-02	RYD
4.61E+01	2.01E+00	1.63E+01	1.16E+01	5.98E+00	1.6E-01	1.08E+00	2.58E+00	3.20E-01	BAS
6.84E+01	7.90E-01	1.25E+01	2.13E+00	1.13E+00	1.5E-01	4.37E+00	3.24E+00	1.50E-01	TAN
6.36E+01	6.20E-01	1.49E+01	3.42E+00	2.89E+00	6.0E-02	4.85E+00	2.27E+00	2.10E-01	DAC
6.41E+01	6.10E-01	1.65E+01	5.62E+00	2.34E+00	1.3E-01	1.00E+00	4.03E+00	1.50E-01	DAC

	AND	BAS	DAC	RYD	RYL	TAN	TRA	TRB
#1	111	6	7	1	0	13	2	0
#2	13	107	13	4	11	14	11	25
#3	9	1	137	16	8	2	3	0
#4	1	0	6	16	7	0	1	0
#5	0	0	6	10	91	1	1	0
#6	12	8	3	1	1	68	15	4
#7	0	0	4	5	3	5	121	1
#8	1	10	0	0	0	11	1	72

F6b

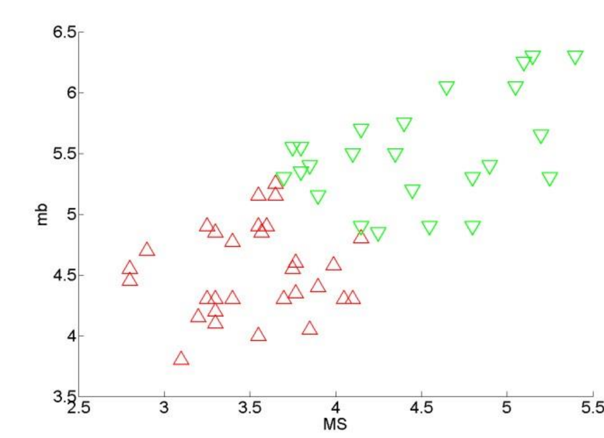
The number of matching classifications is a poor measure of success when classes are not evenly represented. In F6a - classification of infrasound signals – the score is 88%, but 68% can be reached just by a random guess. K-Statistics allows to assess the significance of the classification. Compared to F6a the classification of rocks based on the geochemical composition (F6b) is more significant. K reaches 0.68, higher than the 0.64 in F6a, even though the score is only 72 %. As classes are more evenly represented, the score of random success is only some 15%.



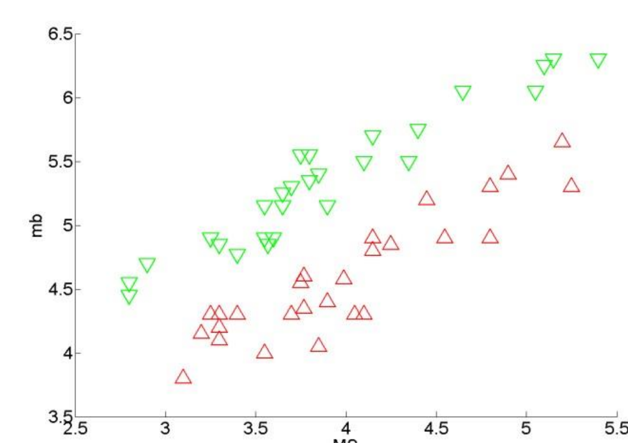
F7

Failure of Supervised Learning provides lessons on the problem, analysis for the reasons is recommended. One reason of failure is improper definition of targets, e. g., the relation of patterns to objects change with time making a-priori information obsolete. In F7 we consider the geochemical composition of rocks erupted from various eruptive centers on Mt Etna, representing it on a Self-Organizing Map. We see that the characteristics of material erupted at a certain center changes with time. This turns out applying **Unsupervised Learning** techniques.

F8a



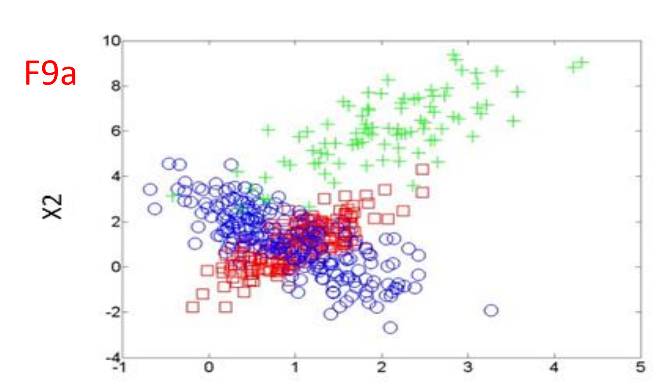
8b



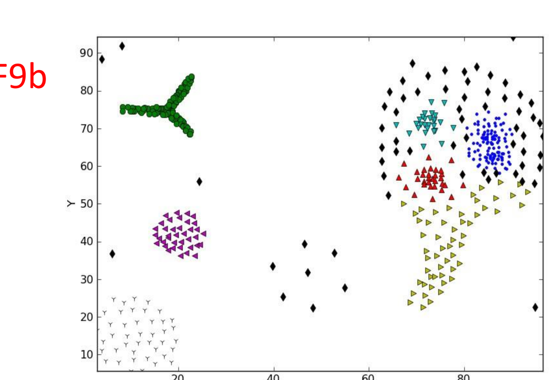
Unsupervised Learning works without a-priori defined targets and is based on **metrics** describing the similarity of patterns among each other. A critical point is the choice of the metric, which depends on the user's ideas. K-means clustering applies the Euclidean metrics (see F8a for the earthquakes-nukes example discussed in F2). In F8b we use a metric which accounts for the fact that the features may be correlated – mb and MS are correlated. The outcome of the clustering depends on the a-priori chosen metrics.

A further issue is the choice of the number of clusters. In K-means clustering formal criteria, e. g., the Davies Bouldin Index (ratio of between- and within-variance), can be used. The criteria need a common metric valid for all clusters. This is not always available (e. g., correlated features where metrics vary among the clusters (F9a). **Density Based Clustering** lacks such, and centroids are not necessarily valid prototypes (F9b, c).

F9a

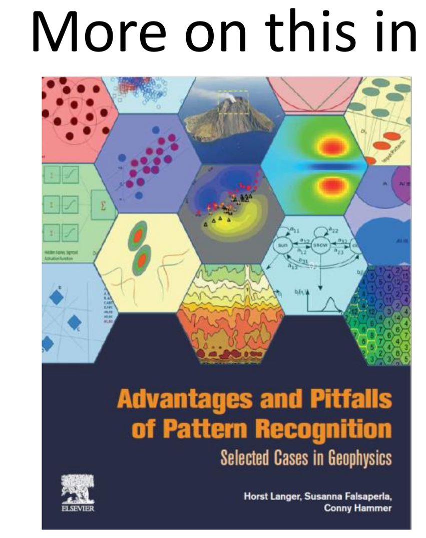


F9b



F9c

More on this in:



Take Home Message

Supervised Learning: Binary Problem : Check TP/FP using ROC & AUC
Multiclass Problem: Verify Significance (e. g., Kappa) !
A-priori Choice of Target can be misleading

Unsuperv. Learning: Data driven and User defined
Metric must be defined by User
Number of clusters often to be defined a-priori
Formal criteria not always applicable
Heuristic criteria, a-posteriori analysis on physical meaning