Introduction

Machine learning field approaches the creation of computer programm's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005). Classification techniques such as Euclidean and Mahalanobis similarity measurements are considered classical machine learning methods. Those similarity measurements are referred to as distance attributes (Marie and Deza, 2016). Euclidean classifiers include a calculation of a centroid on the space of attributes. While the Mahalanobis takes into consideration the shape of attributes space. Both techniques are capable of making identification of geological cyclicity in data.

A Self Organizing Map (SOM) is inspired by neural cortex (Kohonen, 1989). A SOM algorithm is based on a network (Haykin, 2001). This geometric arrangement is an oriented graph, whose vertices are the fundamental units know as artificial neurons and the edges are weights governing the interactions among neurons. Those artificial neurons change their weights as iteractions go on.

This work aims to define a comparison between a Kohonen SOM, an euclidean and a mahalanobean classificators. This comparison uses two well log data from a synthetic syneclises sedimentary basin type. It is remarkable that the Mahalanobis classifier produced a higher error when compared to the Euclidean classifier and the SOM. The SOM presented better results for the two synthetic examples, with an error of 0.7% for the first well and 1.5% for the second. In contrast, Mahalanobis and Euclidean classifiers presented an error of 18.3% and 1.7% respectively for the first well and 11.3% and 6% for the second.

Methodology

In a general overview, the methodology adopted in this work is divided into three main parts. The first generates a synthetic syneclises sedimentary basin in which three synthetic wells are drilled (see Fig. 1). The second part uses well log T1 to train the Kohonen SOM and obtain an optimal distribution of weights. Additionally, the same well is again used to store T1 log data into arrays. The last is to use the three techniques and compare the classified patterns for wells C1 and C2.

Synthetic Sedimentary Basin

The proposed model for the machine learning tests was based on a schematic geological model proposed by Mohriak et al. (2008) for the Solimões Sedimentary Basin, North part of Brazil. This modelling reproduces structures such as Horts, Grabens, normal and reverse faults. Fig. 1 shows the model with a zoom box highlighting the non-parallel contact of two differents lithotypes, where three wells were sampled. Four physical data properties were considered: density, gamma-ray, resistivity and velocity (see Tab. 1). The sample rate for the well data is 0.01 observation/meter with contamination of 5% gaussian noise.

Rock	Density (g/cm^3)	Gamma-ray (Ci/g)	Resistivity (Ω/m)	Velocity (Km/s)
Conglomerate	2.30	100.0	6000	2
Shale	2.55	100.0	1000	3
Dolomite	2.72	8.30	3.5×10^3	6
Diabase	2.91	30.0	15×10^7	5.5
Crystalline	2.80	0.7	1.3×10^6	5

Table 1 *Physical properties.*

Training and Similarities

SOM are machine learning types composed by oriented graphs that are distributed on a hyperplane inside a hyperspace of features. Features are the physical properties that are correlated with a specific type of

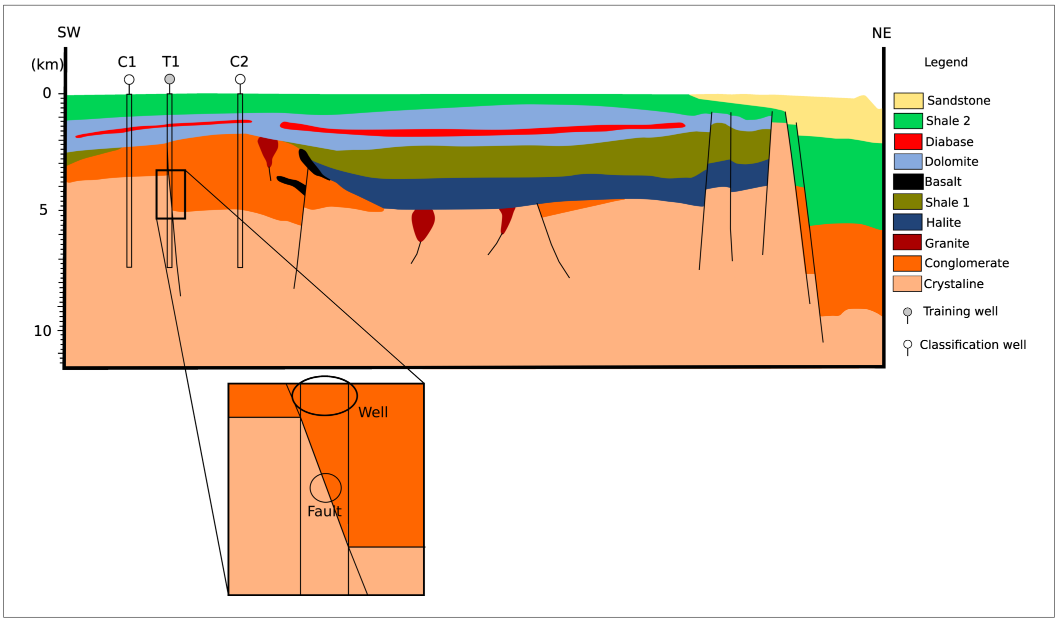

Figure 1 *Synthetic Sedimentary Basin by Mohriak et al. (2008) T1, C1 and C2 are training and classifing wells respectively.*

rock. The identification process is based on redundance of patterns. A toroid geometry is adopted here for the SOM with 400 neurons. Vector **X** is composed of physical properties from the training well T1 with dimension n, where n is the number of data. **X** is related to a neuron with a $w_{i,j}$ weight matrix, as follows:

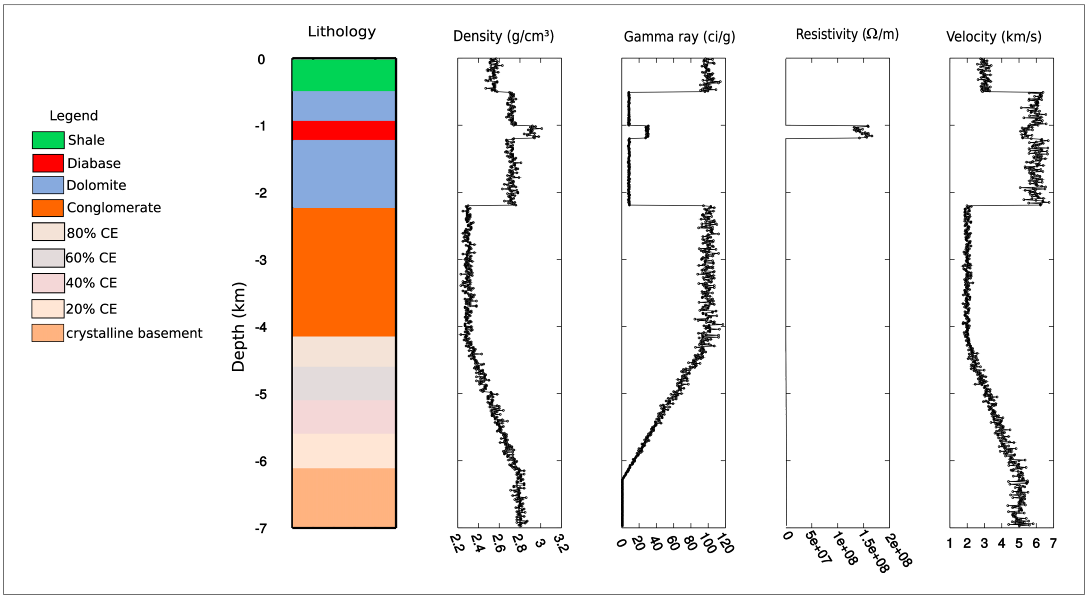

Figure 2 *Synthetic training well $T1$. The normal fault was created by four divisions on the range of the normal fault, decreasing the amount of conglomerate in comparison to cryshalline basement. That behavior simulates a special kind of logging signature.*

$$d(t) = \sqrt{\sum_{i=1}^{n}[x(t) - w_{i,j}(t)]^2} \qquad (j = 1,..,m), \tag{1}$$

where t is the number of iterations, $x(t)$ is an element of $\mathbf{X}$ and $d(t)$ is the metric. The lowest value of $d(t)$ defines the best neuron for a specifc attribute.

The Euclidean classifier is a statistical non-parametric classifier that uses the same input vector $\mathbf{X}$ for its training process. A mean vector $\bar{\mathbf{X}}_i$ for each property is calculated and stored in a set of training i. Then the Euclidian distance for the i-th set (Ed_i) is computed as the following:

$$Ed_i = \|\mathbf{X} - \bar{\mathbf{X}}_i\|^{\frac{1}{2}}, \tag{2}$$

where $\mathbf{X}$ is the set of physical properties to be classified.

The Mahalanobis Classifier computes the mean vector $\bar{\mathbf{X}}_i$ and the covariance matrix $\mathbf{C}_i$ for each ensemble. The Mahalanobis distance (Md_i) is computed as:

$$Md_i = [(\mathbf{X} - \bar{\mathbf{X}}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \bar{\mathbf{X}}_i)]^{\frac{1}{2}}. \tag{3}$$

The covariance matrix is defined as:

$$\mathbf{C}_i = \frac{1}{n_i - 1} \sum_{X \in \omega_i} (\mathbf{X} - \bar{\mathbf{X}}_i)(\mathbf{X} - \bar{\mathbf{X}}_i)^T \tag{4}$$

Results

Fig. 3 (A) shows the original well. Fig. 3 (B), (C) and (D) present the final classification for SOM, Euclidean and Mahalanobis. All errors were concentrated on a single lithotype, the crystalline rock. Those errors indicate 11 swaping between crystalline rock and 20%CE rock type for the SOM classificator.

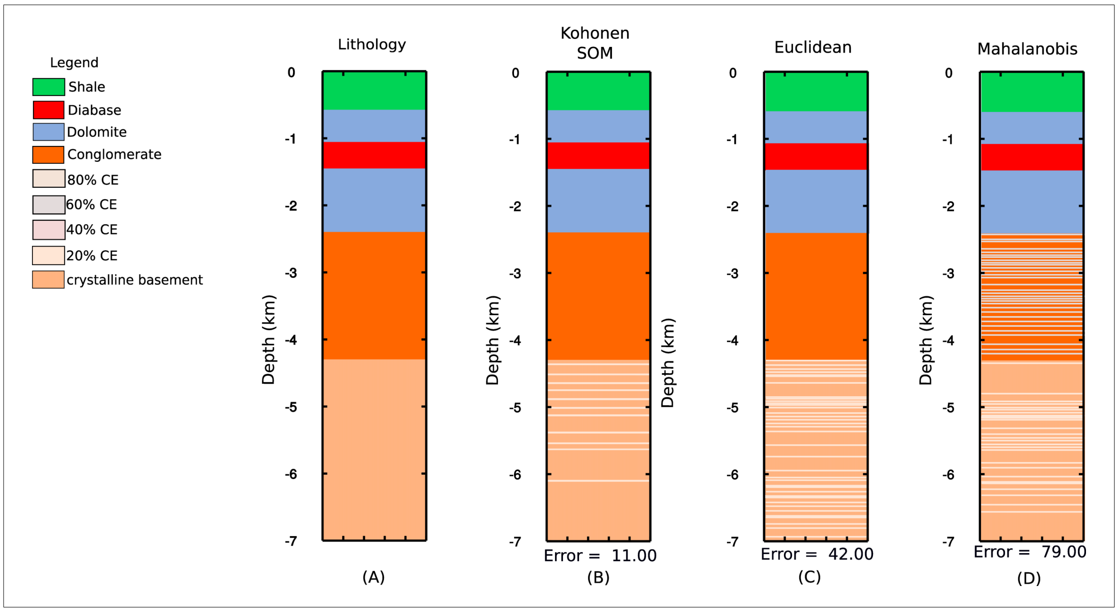

Figure 3 *Comparison between the classificators and SOM for C1 well data.*

Euclidean classificator results on 42 errors. Those errors present the same pattern of SOM classificators. Mahalanobis classificator shows 79 errors. It misjudgeds the crystalline rock data with the 20%CE rock and the conglomerate with the 60%CE rock.

Fig. 4 (A) shows the synthetic well. Fig. 4 (B), (C) and (D) present the comparison among the three methodologies. In a overall perspective, SOM shows better results. Again for the three methods, the major misleading occurs in the classification of crystalline basement.

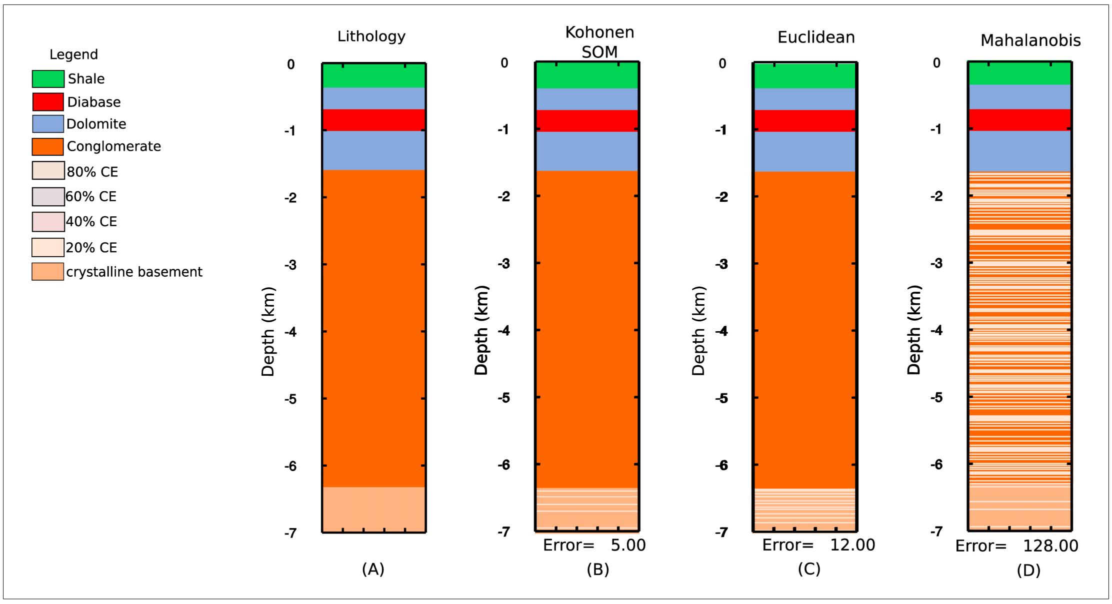

Figure 4 *Comparison between the classificators and SOM for C2 well data.*

Conclusions

Two synthetic tests of well logging were performed to understand the behavior of three different machine learning elements: Kohonen SOM, Euclidean and Mahalanobis classificators. As expected, the SOM overperformed the classificators due to a more detailed algorithm. On the other side, the computational requirements for SOM are more demanding than the classificators, which indicates that the choice of the method depends on the number of data sets.

As perspectives, we are intended to apply these methods to more complex synthetic scenarios and also with real data acquired on Paraná Sedimentary Basin, South portion of Brazil.

References

Haykin, S. [2001] *Neural networks: principles and practice.* McMaster University, Hamilton, Ontario - Canada., 2 edn.

Kohonen, T. [1989] Biological Cybernetics 9 1989. **425**, 139–145.

Levy, S. [1997] The Computer. *Newsweek*, **130**(22), 28.

MacKay, D.J.C. [2005] *Information Theory, Inference, and Learning Algorithms David J.C. MacKay*, 100.

Marie, D.M. and Deza, E. [2016] *Encyclopedia of Distances*. Springer, Moscow State Pedagogical University, 4 edn.

Michie, E.D., Spiegelhalter, D.J. and Taylor, C.C. [1994] Machine Learning , Neural and Statistical Classification. *Technometrics*, **37**(4), 459.

Mohriak, W., Szatmari, P. and Anjos, S. [2008] *Salt: Geology and Tectonic. Exemples on Brazilian's Sedimentary Basins. (In Portuguese)*. Beca, São Paulo, SP., 1 edn.