2018 EAGE ANNUAL
80TH CONFERENCE + EXHIBITION
COPENHAGEN | DENMARK

Abstract No Th K 14

A Comparison of Machine Learning Processes for Classification of Rock Units Using Well Log Data

Victor Carreira[1], Cosme Ponte[1], Rodrigo Bijani[1].

EAGE
EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

[1] Observatório Nacional

11-14 JUNE 2018
WWW.EAGEANNUAL2018.ORG

Changing seasons

Varvite (Itu - São Paulo)

Introduction

- Machine learning approaches computer program's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005).

Introduction

- Machine learning approaches computer program's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005).
- Classification techniques uses distances attributes (Michel and Deza, 2016).

Introduction

- Machine learning approaches computer program's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005).
- Classification techniques uses distances attributes (Michel and Deza, 2016).
- Euclidean classifier calculates a centroid in the space of attributes.

Introduction

- Machine learning approaches computer program's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005).
- Classification techniques uses distances attributes (Michel and Deza, 2016).
- Euclidean classifier calculates a centroid in the space of attributes.
- Mahalanobis classifier takes into consideration the shape of attributes space.

Introduction

- Machine learning approaches computer program's that have the capability of automatically improve themselves through experience (Michie et al., 1994; Levy, 1997; MacKay, 2005).
- Classification techniques uses distances attributes (Michel and Deza, 2016).
- Euclidean classifier calculates a centroid in the space of attributes.
- Mahalanobis classifier takes into consideration the shape of attributes space.
- A Self Organizing Map (SOM) is inspired by neural cortex (Kohonen, 1989) and based oriented graph (Haykin, 1999) working as an interconnected network.

Objective

- Define comparisons between a Kohonen (SOM), an euclidean and a mahalanobean classifiers.

Objective

- Define comparisons between a Kohonen (SOM), an euclidean and a mahalanobean classifiers.
- Using synthetic well log data in a controlled situation.

Methodology

Compare the results

Uses C1 and C2 to
make predictions using
SOM and the classifiers

Uses T1 well data to train the
Kohonen (SOM) and make
the statistical of classifiers

Generates the hypothesis model

Synthetic Sedimentary Basin

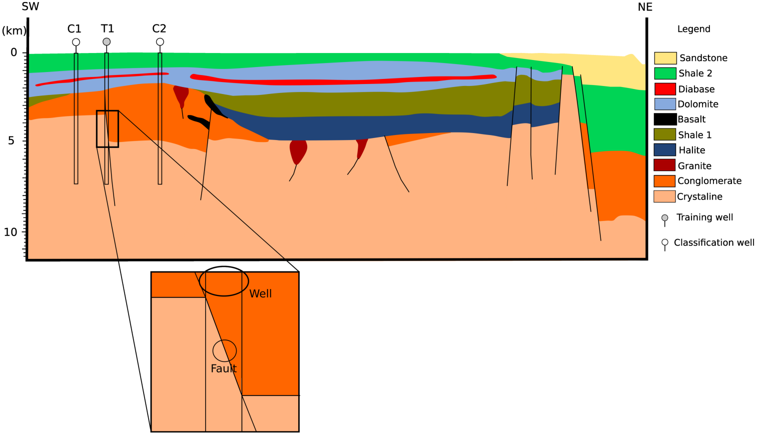


Figure : Synthetic Sedimentary Basin by Mohriak et al. (2008) T1, C1 and C2 are training and classifing wells respectively.

Synthetic Sedimentary Basin

Table : Physical properties and rock types.

Rock	Density (g/cm^3)	Gamma ray (Ci/g)	Resistivity ($\Omega.m$)	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

- The sample rate for the well data is 0.01 observation/meter with contamination of 5% Gaussian noise.

Training and Similarities

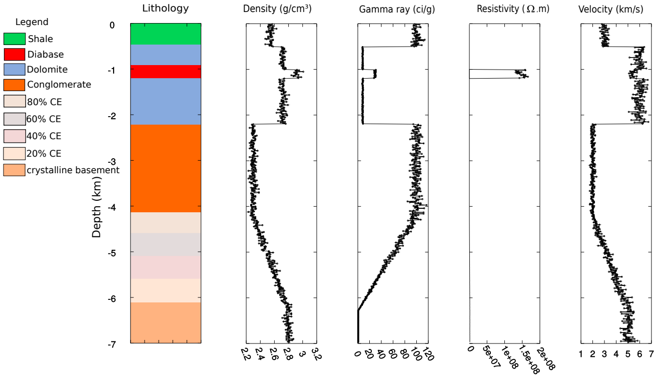


Figure : Synthetic training well T1.

Training and Similarities

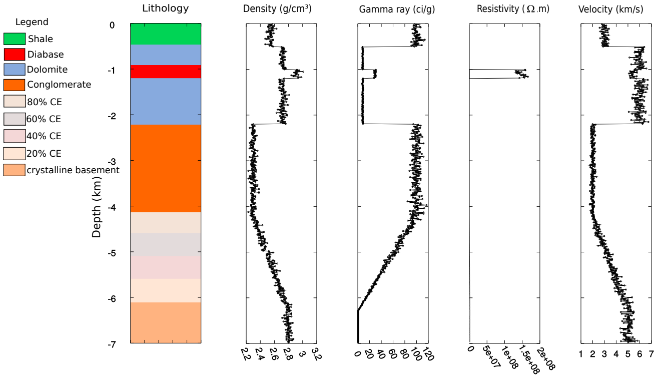


Figure : Synthetic training well T1.

- Four divisions describes the normal fault by decreasing the amount of conglomerate in comparison to crystalline basement.

Training and Similarities

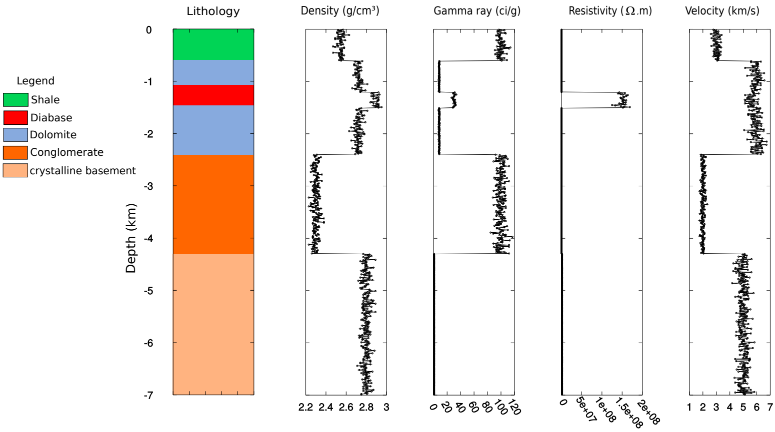


Figure : Classification well C1

Training and Similarities

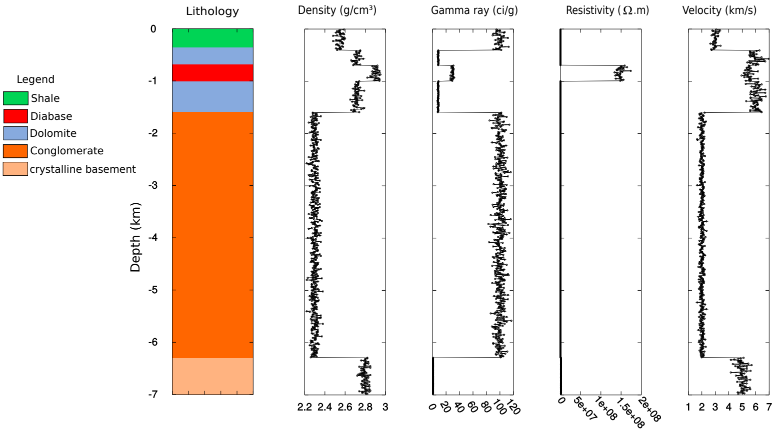


Figure : Classification well C2

Euclidean Classifier

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

Euclidean Classifier

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

$\mathbf{X}$, input vector

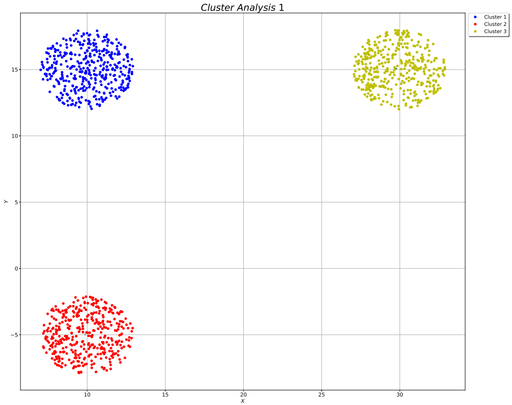
Euclidean Classifier

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

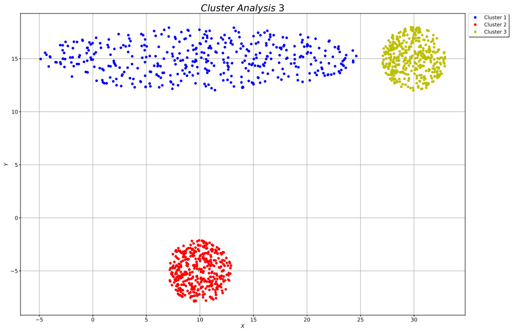
$\mathbf{X}$, input vector

$\bar{\mathbf{X}}_i$, mean vector

Euclidean Classifier



Euclidean Classifier



Mahalanobis Classifier

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

Mahalanobis Classifier

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

X , input vector

Mahalanobis Classifier

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

$\mathbf{X}$, input vector

$\bar{\mathbf{X}}_i$, mean vector

THE EAGE ANNUAL 2018

Mahalanobis Classifier

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

$\mathbf{X}$, input vector

$\bar{\mathbf{X}}_i$, mean vector

$\mathbf{C}_i$, covariance matrix

Mahalanobis Classifier

$$\mathbf{C}_i = \frac{1}{n_i - 1} \sum_{X \in \omega_i} (\mathbf{X} - \bar{\mathbf{X}}_i)(\mathbf{X} - \bar{\mathbf{X}}_i)^T \qquad (3)$$

Mahalanobis Classifier

$$\mathbf{C}_i = \frac{1}{n_i - 1} \sum_{X \in \omega_i} (\mathbf{X} - \bar{\mathbf{X}}_i)(\mathbf{X} - \bar{\mathbf{X}}_i)^T \tag{3}$$

n_i , number of elements

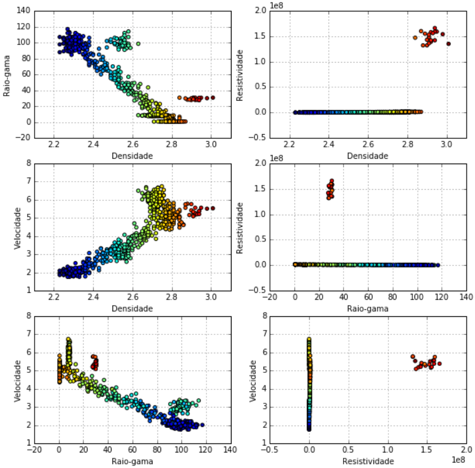
Mahalanobis Classifier

$$\mathbf{C}_i = \frac{1}{n_i - 1} \sum_{X \in \omega_i} (\mathbf{X} - \bar{\mathbf{X}}_i)(\mathbf{X} - \bar{\mathbf{X}}_i)^T \tag{3}$$

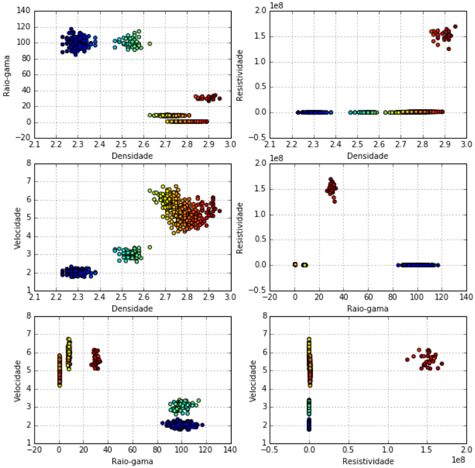
n_i , number of elements

ω_i , space of attributes

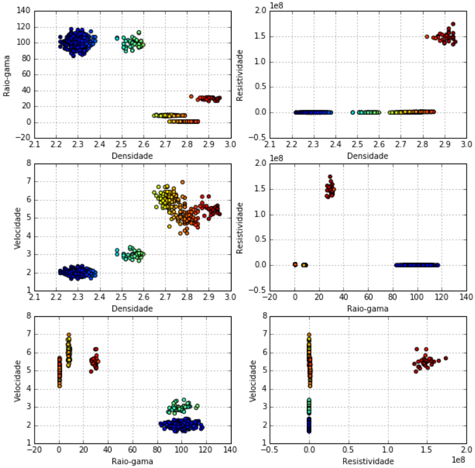
Clusters and Space of Attributes - T1 well



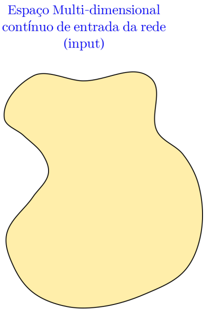
Clusters and Space of Attributes - C1 well



Clusters and Space of Attributes - C2 well

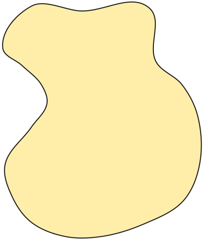


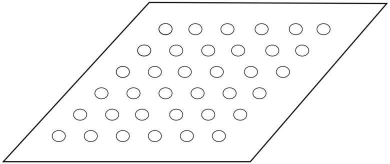
Kohonen - SOM


Espaço Multi-dimensional
contínuo de entrada da rede
(input)

Kohonen - SOM

Espaço Multi-dimensional
contínuo de entrada da rede
(input)

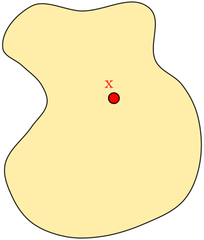


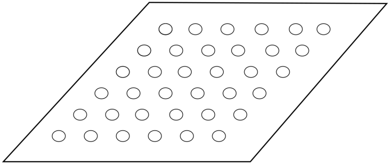


Espaço Uni-dimensional
discreto de saída da rede
(output)

Kohonen - SOM

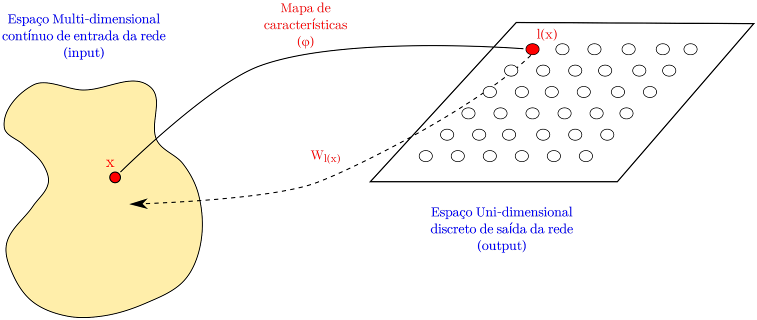
Espaço Multi-dimensional
contínuo de entrada da rede
(input)





Espaço Uni-dimensional
discreto de saída da rede
(output)

Kohonen - SOM


Espaço Multi-dimensional
contínuo de entrada da rede
(input)
Mapa de
características
(φ)
l(x)
x
$W_{l(x)}$
Espaço Uni-dimensional
discreto de saída da rede
(output)

Kohonen - SOM

Organization

$$\mathbf{x} = [x_1, x_2, x_3, \ldots, x_m]^T$$

Kohonen - SOM

Organization

$$\mathbf{x} = \left[x_1, x_2, x_3, \dots, x_m\right]^T$$

$$\mathbf{w}_j = \left[w_{j1}, w_{j2}, w_{j3}, \dots, w_{jm}\right]^T$$

Kohonen - SOM

Organization

$$\mathbf{x} = [x_1, x_2, x_3, ..., x_m]^T$$

$$\mathbf{w}_j = [w_{j1}, w_{j2}, w_{j3}, ..., w_{jm}]^T$$

$$j = 1, 2, 3, \ldots, l$$

Kohonen - SOM

Cooperation

$$i(\mathbf{x}) = argmin_j \parallel \mathbf{x} - \mathbf{w}_j \parallel$$

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

d(t) distance or identity of a neuron i

Kohonen - SOM

Synaptic adaption or Training process

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)[x(t) - w_{i,j}(t)]$$

Kohonen - SOM

Synaptic adaption or Training process

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)[x(t) - w_{i,j}(t)]$$

$w_{i,j}(t+1)$, updated weight matrix

$\eta(t)$, learning rate

$$\eta(t) = \eta(0)(1 - \frac{t}{T})$$

T , number of training cyles

Kohonen - SOM

Synaptic adaption or Training process

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)[x(t) - w_{i,j}(t)]$$

$w_{i,j}(t+1)$, updated weight matrix

$\eta(t)$, learning rate

$$\eta(t) = \eta(0)(1 - \frac{t}{T})$$

T , number of training cyles

t , number of iteractions

Kohonen - SOM

Synaptic adaption or Training process

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta(t)[x(t) - w_{i,j}(t)]$$

$w_{i,j}(t+1)$, updated weight matrix

$\eta(t)$, learning rate

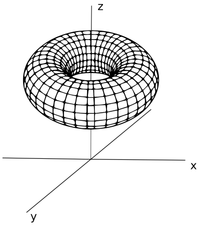
$$\eta(t) = \eta(0)(1 - \frac{t}{T})$$

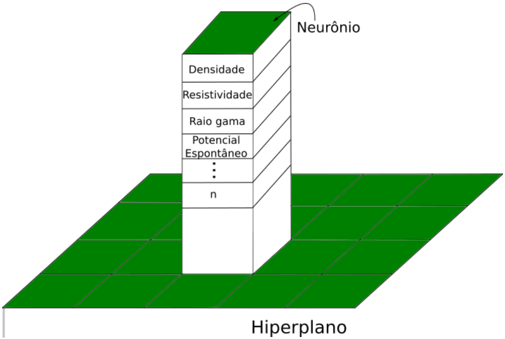
T , number of training cyles

t , number of iteractions

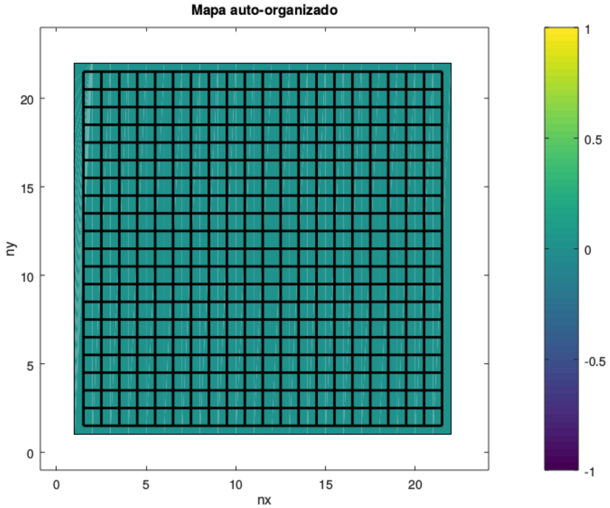
- A iteractive process $t = t + 1$ goes on until $t \approx T$. Once the process ends for one neuron it repeats it self for the surrounding neighbors (YANG et al., 2009; Yan et al., 2014).

The geometry



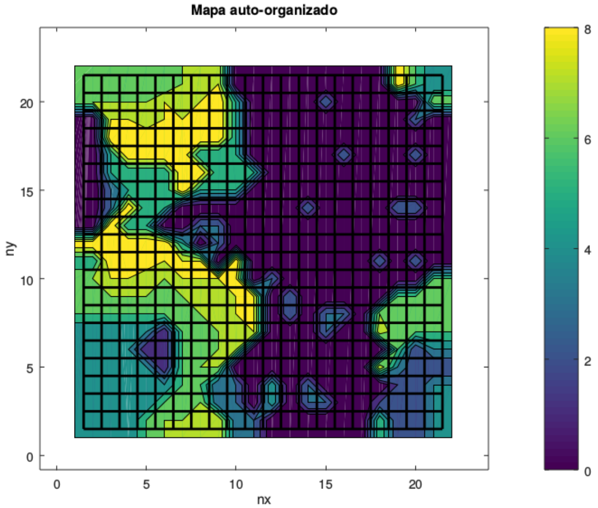


A hyperplane with 400 neurons



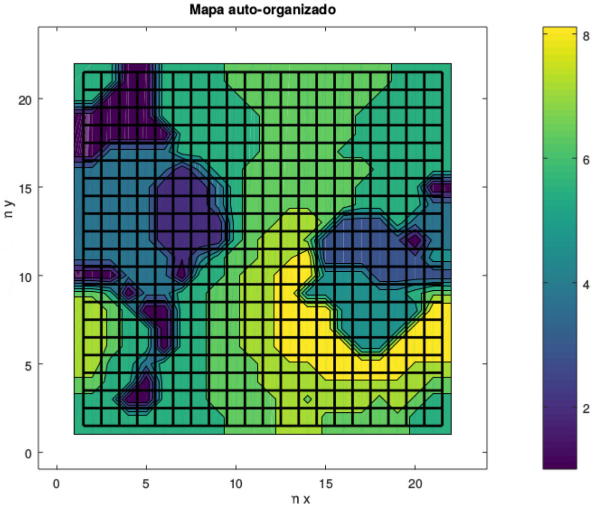
THE EAGE ANNUAL 2018

A hyperplane with 400 neurons



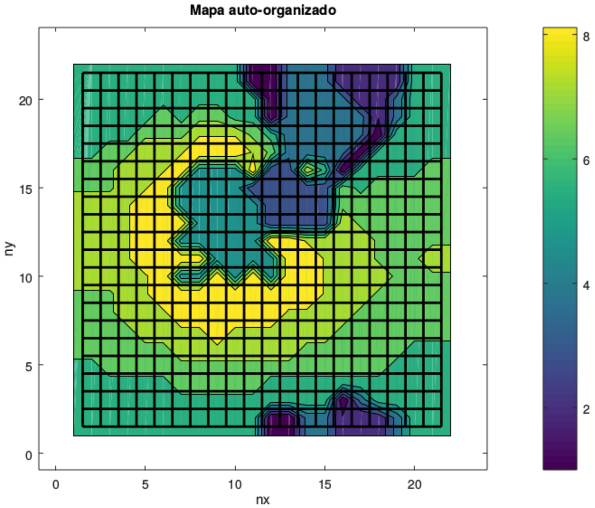
THE EAGE ANNUAL 2018

A hyperplane with 400 neurons



THE EAGE ANNUAL 2018

A hyperplane with 400 neurons

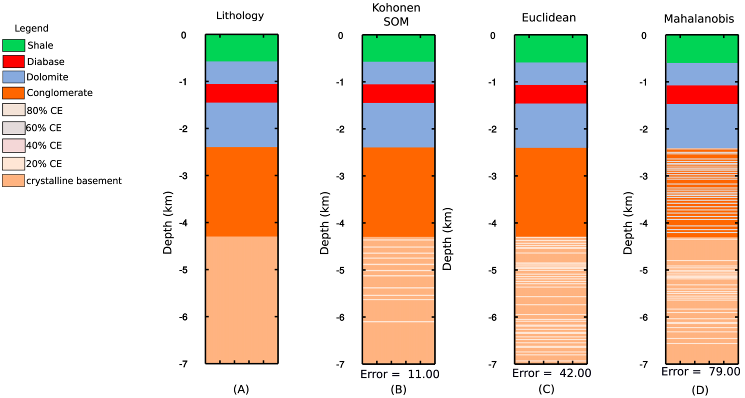


A hyperplane with 400 neurons

Lithology	Normalization
Shale 2	1
Dolomite	2
Diabase	3
Conglomerate	4
Conglomerate 80%	5
Conglomerate 60%	6
Conglomerate 40%	7
Conglomerate 20%	8
Crystalline Basement	9

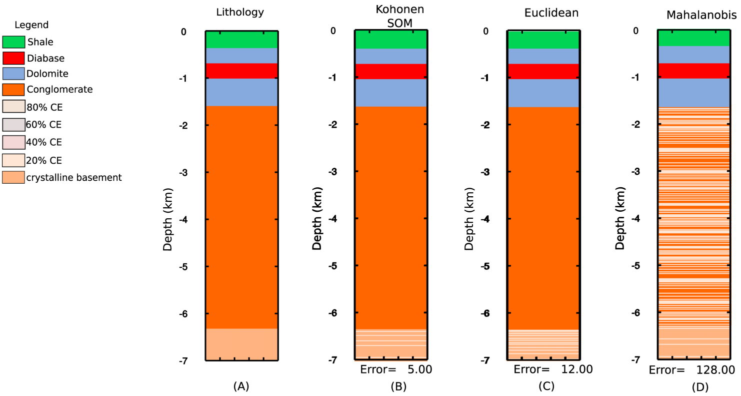
Results

Classifications C1



Results

Classifications C2



THE EAGE ANNUAL 2018

Conclusions

- SOM algorithm overperformed the euclidean and mahalanobis classifiers with an error of 0.71%

Conclusions

- SOM algorithm overperformed the euclidean and mahalanobis classifiers with an error of 0.71%
- The complex distribution of features on the space of properties led to an increase of errors on classifiers of 18.28% (mahalanobis) and 1.74% (euclidean)

Conclusions

- SOM algorithm overperformed the euclidean and mahalanobis classifiers with an error of 0.71%
- The complex distribution of features on the space of properties led to an increase of errors on classifiers of 18.28% (mahalanobis) and 1.74% (euclidean)
- Classifiers can not perform classification of bell patterns in well data.

Conclusions

- SOM algorithm overperformed the euclidean and mahalanobis classifiers with an error of 0.71%
- The complex distribution of features on the space of properties led to an increase of errors on classifiers of 18.28% (mahalanobis) and 1.74% (euclidean)
- Classifiers can not perform classification of bell patterns in well data.
- Box patterns could be solved with euclidean classifier but not in mahalanobis

Conclusions

- SOM algorithm overperformed the euclidean and mahalanobis classifiers with an error of 0.71%
- The complex distribution of features on the space of properties led to an increase of errors on classifiers of 18.28% (mahalanobis) and 1.74% (euclidean)
- Classifiers can not perform classification of bell patterns in well data.
- Box patterns could be solved with euclidean classifier but not in mahalanobis
- Kohonen - SOM showed the bests results concerning classification of mixture of rocks or the bell pattern

References I

Haykin, S. (1999). *Neural Networks: a comprehensive foundation*. Prentice Hall Inc., Porto Alegre, 2 edition.

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*, volume 100.

Michel, M. D. and Deza, E. (2016). *Encyclopedia of Distances*. 4 edition.

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). *Sal: Geologia e Tectônica. Exemplos nas Bacias Brasileiras*. Beca, São Paulo, SP., 1 edition.

Yan, Z., Xiaodong, Z., Jiaotong, L., and Corporation, G. (2014). Lithology identification research based on self-organizing map of data mining method.

YANG, C. T., MARSOOLI, R., and AALAMI, M. T. (2009). Evaluation of total load sediment transport formulas using ANN. *International Journal of Sediment Research*, 24(3):274–286.

Questions?

Author: Victor Carreira
E-mail: victorcarreira@on.br