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

Observatório Nacional[1]

Changing seasons

Varvite (Itu - São Paulo)

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Machine learning groups:

1. Artifitial Neural Networks (Minsky and Papert, 1969; Michie et al., 1994; MacKay, 2005);
2. Decision trees (Mitchell, 1997; Simard et al., 2000; Roberts et al., 2002);
3. Statistical Classifiers (Michel and Deza, 2016);
4. Self-Organizing Maps (Kohonen, 1989; Haykin, 1999);

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Machine learning groups:

1. Artifitial Neural Networks (Minsky and Papert, 1969; Michie et al., 1994; MacKay, 2005);
2. Decision trees (Mitchell, 1997; Simard et al., 2000; Roberts et al., 2002);
3. Statistical Classifiers (Michel and Deza, 2016);
4. Self-Organizing Maps (Kohonen, 1989; Haykin, 1999);

Classifiers: uses the concept of distance in the space of attributes;

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Machine learning groups:

1. Artifitial Neural Networks (Minsky and Papert, 1969; Michie et al., 1994; MacKay, 2005);
2. Decision trees (Mitchell, 1997; Simard et al., 2000; Roberts et al., 2002);
3. Statistical Classifiers (Michel and Deza, 2016);
4. Self-Organizing Maps (Kohonen, 1989; Haykin, 1999);

Classifiers: uses the concept of distance in the space of attributes;

Euclidean classifier: calculates a centroid in the space of attributes.

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Machine learning groups:

1. Artifitial Neural Networks (Minsky and Papert, 1969; Michie et al., 1994; MacKay, 2005);
2. Decision trees (Mitchell, 1997; Simard et al., 2000; Roberts et al., 2002);
3. Statistical Classifiers (Michel and Deza, 2016);
4. Self-Organizing Maps (Kohonen, 1989; Haykin, 1999);

Classifiers: uses the concept of distance in the space of attributes;

Euclidean classifier: calculates a centroid in the space of attributes.

Mahalanobis classifier: takes into consideration the shape of attributes space.

Introduction

Machine learning: computer program with capability of automatic improvement trought experience.

Machine learning groups:

1. Artifitial Neural Networks (Minsky and Papert, 1969; Michie et al., 1994; MacKay, 2005);
2. Decision trees (Mitchell, 1997; Simard et al., 2000; Roberts et al., 2002);
3. Statistical Classifiers (Michel and Deza, 2016);
4. Self-Organizing Maps (Kohonen, 1989; Haykin, 1999);

Classifiers: uses the concept of distance in the space of attributes;

Euclidean classifier: calculates a centroid in the space of attributes.

Mahalanobis classifier: takes into consideration the shape of attributes space.

Self-Organizing Map: inspired by neural cortex and based oriented graph that works as an interconnected network.

Objective of this work

Geophysical problem: Identify rocks from well log data by means of machine learning and statisticas classifiers;

Objective of this work

Geosphysical problem: Identify rocks from well log data by means of machine learning and statisticas classifiers;

Way to solve it: use self-organizing map (SOM) and two statistical classifers;

Objective of this work

Geosphysical problem:	Identify rocks from well log data by means of machine learning and statisticas classifiers;
Way to solve it:	use self-organizing map (SOM) and two statistical classifers;
Strategy:	Compare the results for the three explored methods, Kohonen (SOM), euclidean and mahalanobean classifiers for a synthetic scenarium.

Synthetic Sedimentary Basin

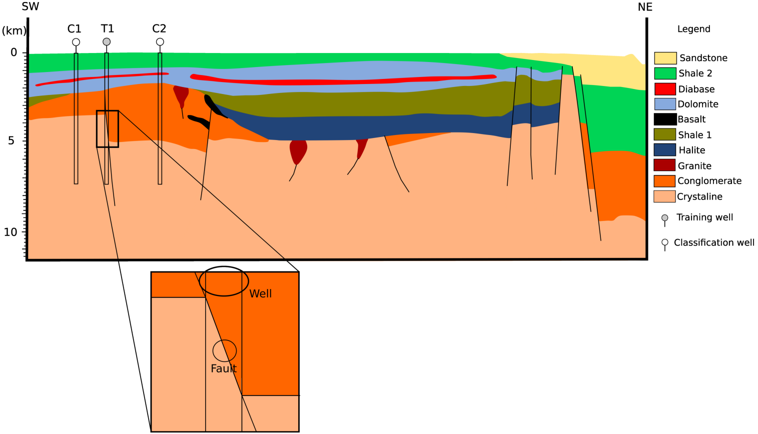


Figure : Synthetic Sedimentary Basin by Mohriak et al. (2008) T1, C1 and C2 are training and classifing wells respectively.

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

Sample rate: 0.01 observation/meter

Contamination error: 5% Radom Gaussian noise.

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

Training and Similarities

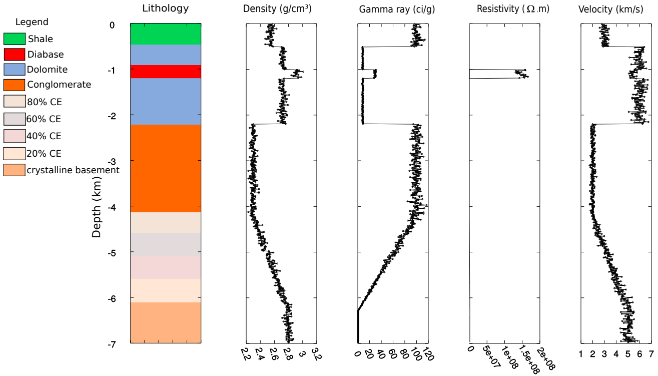


Figure : Synthetic training well T1.

Training and Similarities

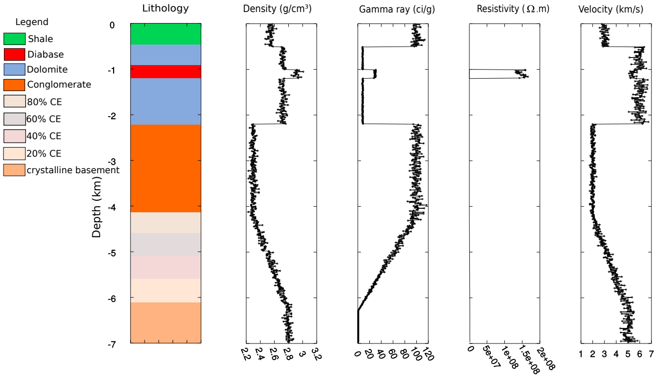


Figure : Synthetic training well T1.

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

Training and Similarities

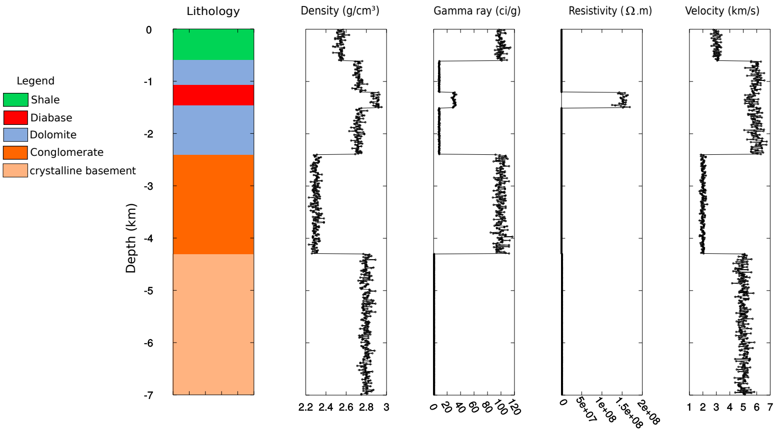


Figure : Classification well C1

Training and Similarities

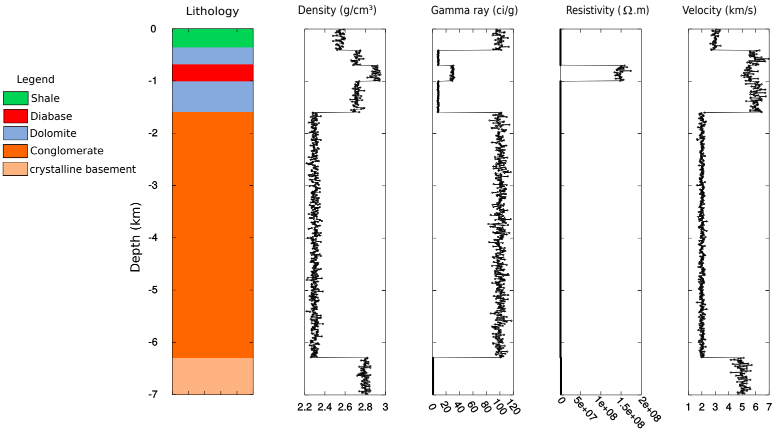


Figure : Classification well C2

Euclidean Classifier

Definition

$$Ed_i = \|\mathbf{X} - \bar{\mathbf{X}}_i\|_2$$

Euclidean Classifier

Definition

$$Ed_i = \|\mathbf{X} - \bar{\mathbf{X}}_i\|_2$$

$\mathbf{X}$, input vector (attribute data)

Euclidean Classifier

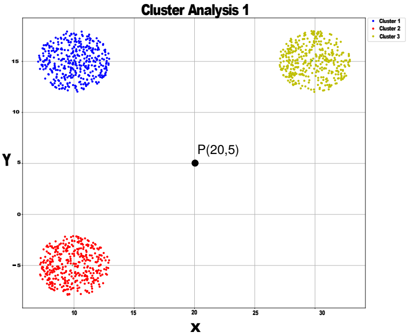
Definition

$$Ed_i = \|\mathbf{X} - \bar{\mathbf{X}}_i\|_2$$

$\mathbf{X}$, input vector (attribute data)

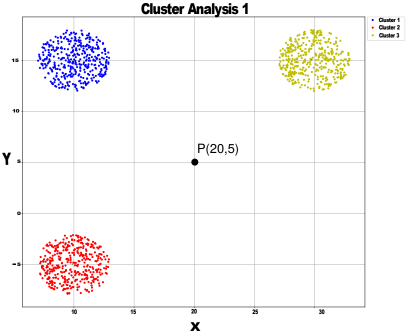
$\bar{\mathbf{X}}_i$, mean vector

Euclidean Classifier



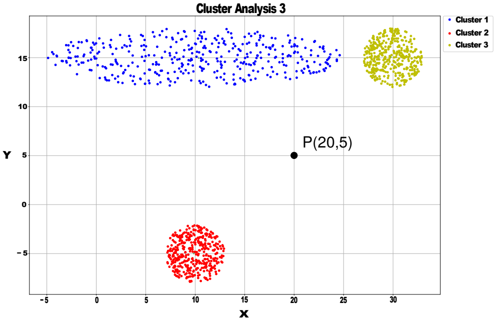
- All centroid are equidistant

Euclidean Classifier



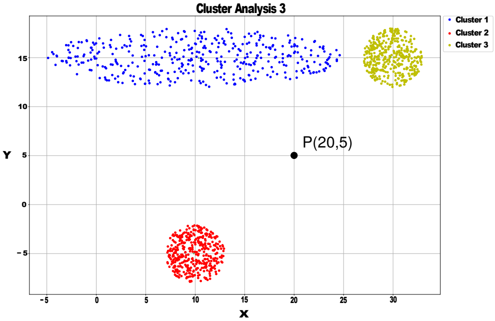
- All centroid are equidistant
- $P(20, 5)$ could be a member of all groups from a euclidean point of view

Euclidean Classifier



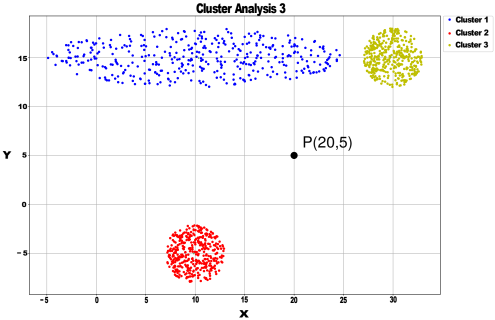
- Blue cluster still have the same centroid

Euclidean Classifier



- Blue cluster still have the same centroid
- Problematic to define which cluster point $(20, 5)$ should be member of

Euclidean Classifier



- Blue cluster still have the same centroid
- Problematic to define which cluster point $(20, 5)$ should be member of
- How could we handle this problem?

Mahalanobis Classifier

Definition

$$Md_i = \|(\mathbf{X} - \bar{\mathbf{X}}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \bar{\mathbf{X}}_i)\|_2$$

Mahalanobis Classifier

Definition

$$Md_i = \|(\mathbf{X} - \bar{\mathbf{X}}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \bar{\mathbf{X}}_i)\|_2$$

$\mathbf{X}$, input vector

Mahalanobis Classifier

Definition

$$Md_i = \|(\mathbf{X} - \bar{\mathbf{X}}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \bar{\mathbf{X}}_i)\|_2$$

$\mathbf{X}$, input vector

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

Mahalanobis Classifier

Definition

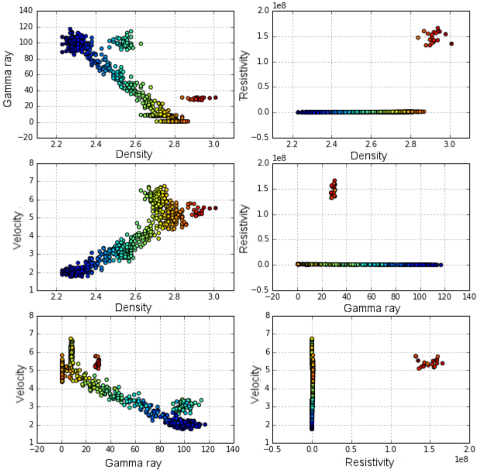
$$Md_i = \|(\mathbf{X} - \bar{\mathbf{X}}_i)^T \mathbf{C}_i^{-1} (\mathbf{X} - \bar{\mathbf{X}}_i)\|_2$$

$\mathbf{X}$, input vector

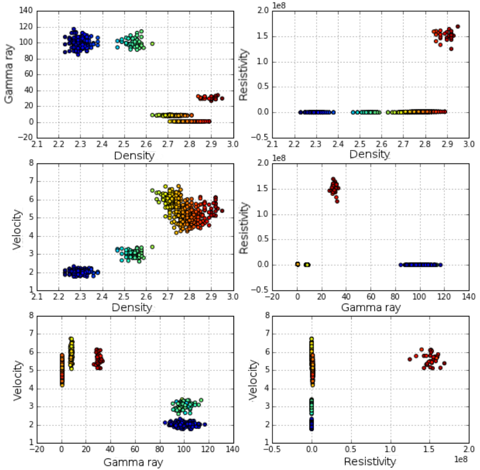
$\bar{\mathbf{X}}_i$, mean vector

$\mathbf{C}_i$, covariance matrix

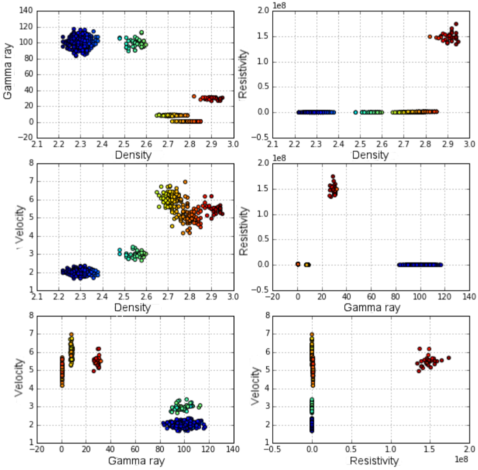
Clusters and Space of Attributes - T1 well



Clusters and Space of Attributes - C1 well

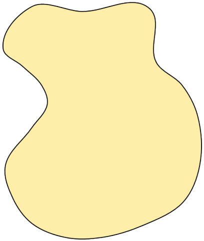


Clusters and Space of Attributes - C2 well



Kohonen - SOM

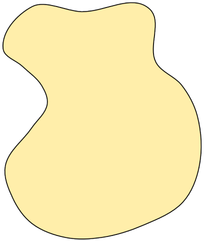
Multi-dimension space of
properties
(input)

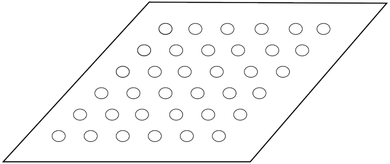


Kohonen - SOM

Sefl Organizing Map (SOM)

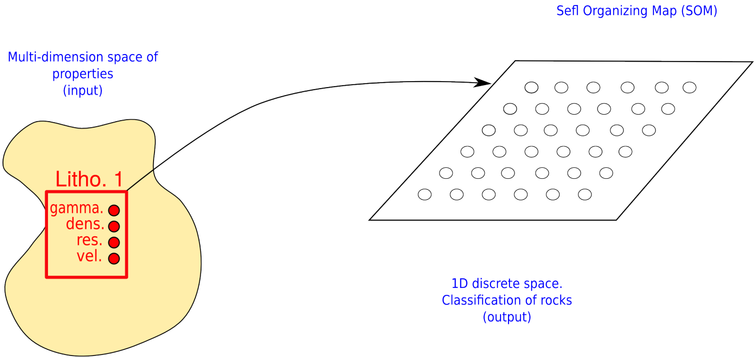
Multi-dimension space of properties (input)





1D discrete space.
Classification of rocks
(output)

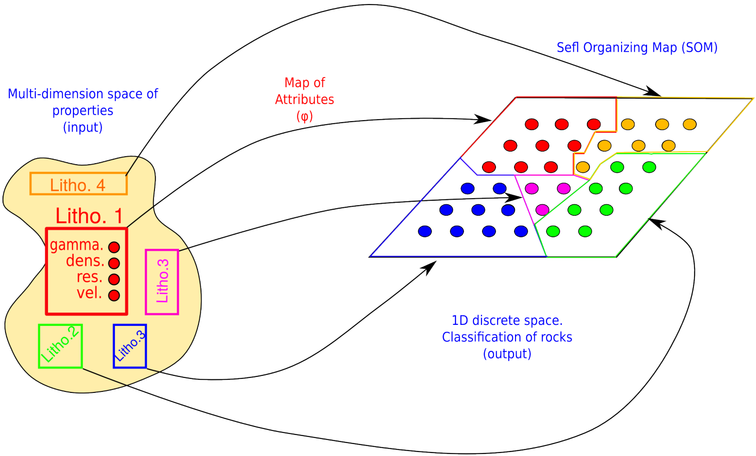
Kohonen - SOM



Lithology = F(gamma.,dens., res., vel.)

Kohonen - SOM

Training



Lithology = F(gamma.,dens., res., vel.)

Kohonen - SOM

Organization

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

Kohonen - SOM

Organization

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

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

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 and winner neuron

$$i(\mathbf{x}) = argmin_j \parallel \mathbf{x} - \mathbf{w}_j \parallel_2$$

i(t) distance or identity of a neuron i

Kohonen - SOM

Synaptic adaption or Training process

Definition

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

Kohonen - SOM

Synaptic adaption or Training process

Definition

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

$w_{i,j}(t+1)$, updated attribute matrix of neurons

$\eta(t)$, learning rate

Kohonen - SOM

Synaptic adaption or Training process

Definition

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

$w_{i,j}(t+1)$, updated attribute matrix of neurons

$\eta(t)$, learning rate

Definition

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

T , number of training cyles

Kohonen - SOM

Synaptic adaption or Training process

Definition

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

$w_{i,j}(t+1)$, updated attribute matrix of neurons

$\eta(t)$, learning rate

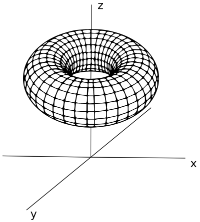
Definition

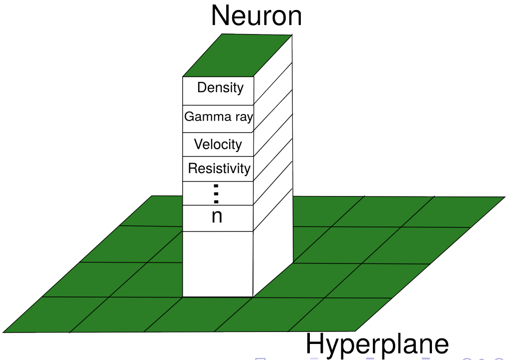
$$\eta(t) = \eta(0)(1 - \frac{t}{T})$$

T , number of training cyles

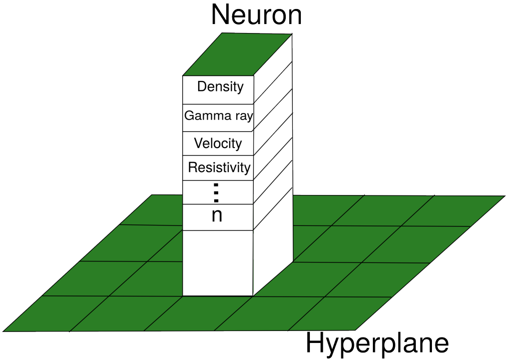
t , number of iteractions

The geometry

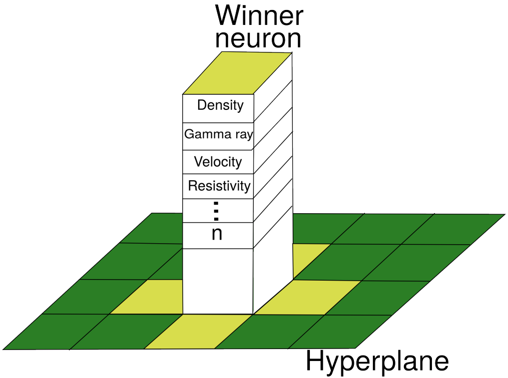




Winner neuron and neighborhood

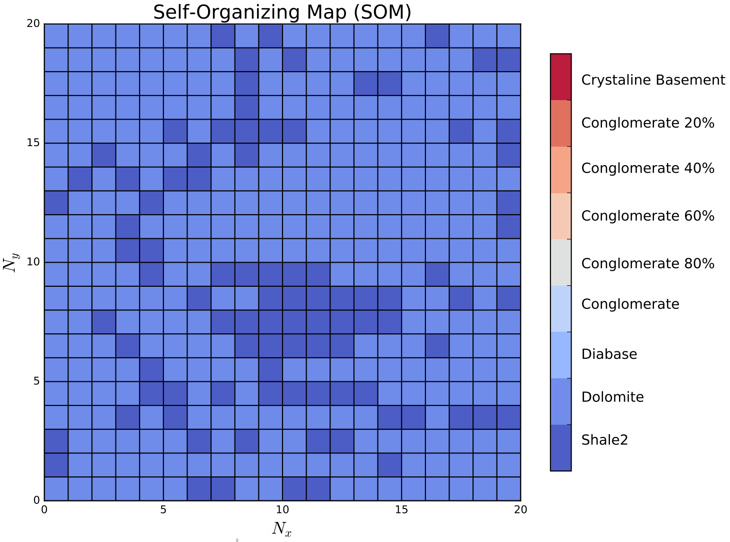


Winner neuron and neighborhood



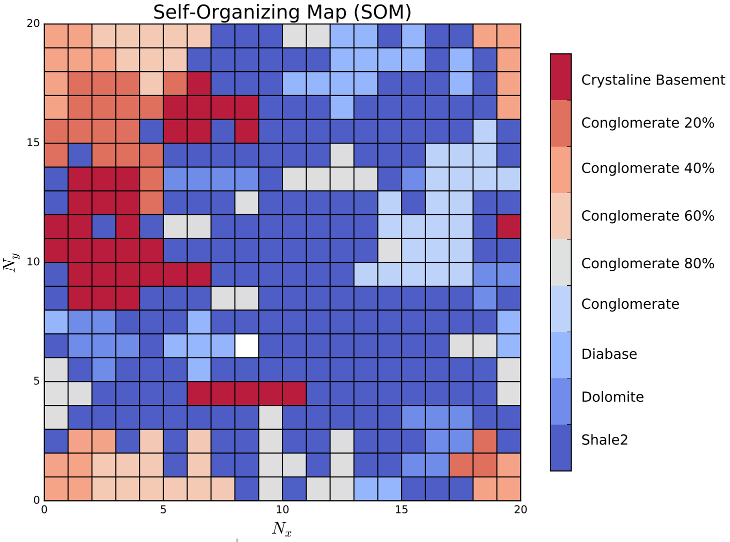
A hyperplane with 400 neurons

Epoch 5



A hyperplane with 400 neurons

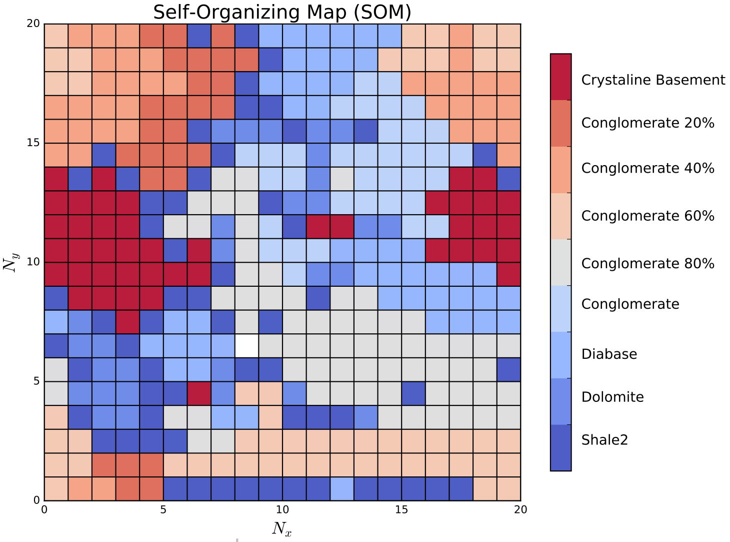
Epoch 100



THE EAGE ANNUAL 2018

A hyperplane with 400 neurons

Epoch 1000



Results

Classifications C1

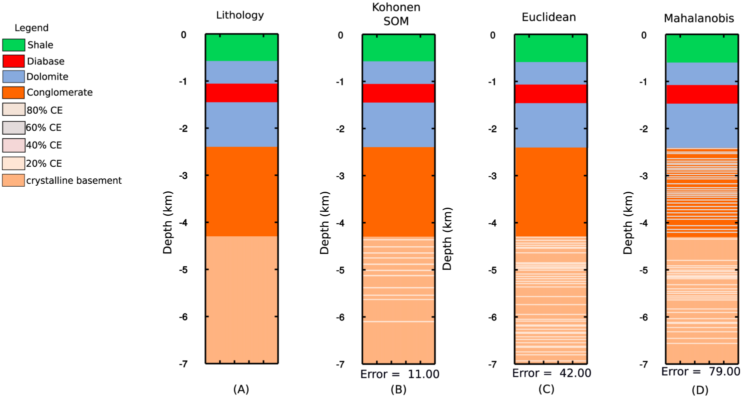


Figure : 700 data analyzed in each well

Results

Classifications C2

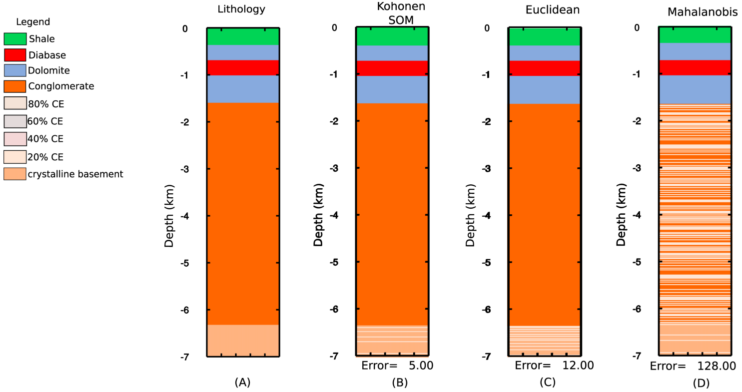


Figure : 700 data analyzed in each well

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

Future Perspectives

Future Perspectives

- Make more tests

Future Perspectives

- Make more tests
- Apply methodology on real data

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.

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.

Minsky, M. and Papert, S. (1969). *Perceptrons*. 2 edition.

Mitchell, T. M. (1997). Decision Tree Learning.

Mohriak, W., Szatmari, P., and Anjos, S. (2008). *Sal: Geologia e Tectônica. Exemplos nas Bacias Brasileiras*. Beca, São Paulo, SP., 1 edition.

Roberts, D. A., Numata, I., Holmes, K., Batista, G., Krug, T., Monteiro, A., Powell, B., and Chadwick, O. A. (2002). Large area mapping of land-cover change in Rond??nia using multitemporal spectral mixture analysis and decision tree classifiers. *Journal of Geophysical Research D: Atmospheres*, 107(20).

Simard, M., Saatchi, S. S., and De Grandi, G. (2000). The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5 I):2310–2321.

Questions?

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