

C. Arouri, N. Bekono, J.C. Fopa, R. Shayakhmetov, S. Aridhi, G. Loosli, C. Roucelle, N. Tsopzé, **E. Mephu Nguifo**

Motivations : Photo-Z redshift reconstruction

MNRAS **000**, 000–000 (0000)

Preprint 1 February 2017

Compiled using MNRAS L^AT_EX style file v3.0

On the realistic validation of photometric redshifts, or why Teddy will never be Happy

R. Beck^{1*}, C.-A. Lin^{2,3}, E. E. O. Ishida⁴, F. Gieseke⁵, R. S. de Souza^{6,7},
M. V. Costa-Duarte^{7,8}, M. W. Hattab⁹, A. Krone-Martins¹⁰,
for the COIN Collaboration

¹*Department of Physics of Complex Systems, Eötvös Loránd University, Budapest 1117, Hungary*

²*Service d'Astrophysique, CEA Saclay, Orme des Merisiers, Bât 709, 91191 Gif-sur-Yvette, France*

³*Fenglin Veteran Hospital, 2 Zhongzheng Rd. Sec. 1, Fenglin Township, Hualien 97544, Taiwan*

⁴*Laboratoire de Physique Corpusculaire, Université Clermont-Auvergne, 4 Avenue Blaise Pascal, 63178, Aubière Cedex, France*

⁵*University of Copenhagen, Sigurdsgade 41, 2200 Copenhagen, Denmark*

⁶*MTA Eötvös University, EIRSA “Lendület” Astrophysics Research Group, Budapest 1117, Hungary*

⁷*Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, R. do Matão 1226, 05508-090, SP, Brazil*

⁸*Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands.*

⁹*Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, VA, USA*

¹⁰*CENTRA/SIM, Faculdade de Ciências, Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016, Lisboa, Portugal*

■ Challenge ?

- ❑ *Lack of spectroscopic coverage in feature space (e.g. colours and magnitudes)*
- ❑ *Mismatch between photometric error distributions associated with the spectroscopic and photometric samples.*

Motivations : Photo-Z redshift reconstruction

- *Regression task*

- Techniques :

- Decision tree, IBL, Bayesian networks, Lattice-based,
 - SVM, Artificial Neural Networks (ANN), ...

- Different models of ANN:

- ...

- **Multilayer perceptron**

- ...

- *Choice of the ANN structure ?*

- *Application to photo-Z redshift reconstruction*

Motivations

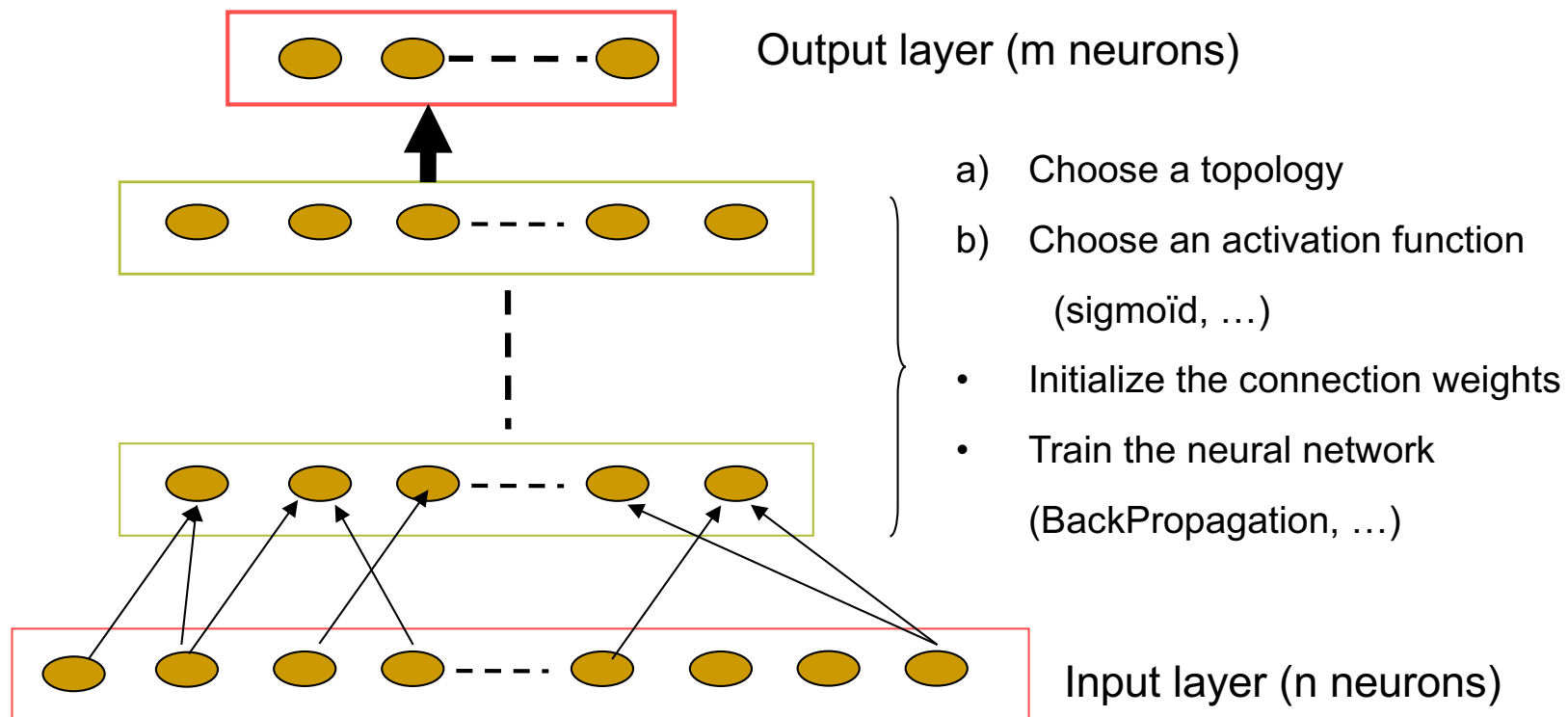
This story might be apocryphal, but it doesn't really matter. It is a perfect illustration of the biggest problem behind neural networks. **Any automatically trained net with more than a few dozen neurons is virtually impossible to analyze and understand.** One can't tell if a net has memorized inputs, or is 'cheating' in some other way. A promising use for neural nets these days is to predict the stock market. *Even though initial results are extremely good, investors are leery of trusting their money to a system that **nobody** understands*

Neil Fraser [2003]

Problem : Settings

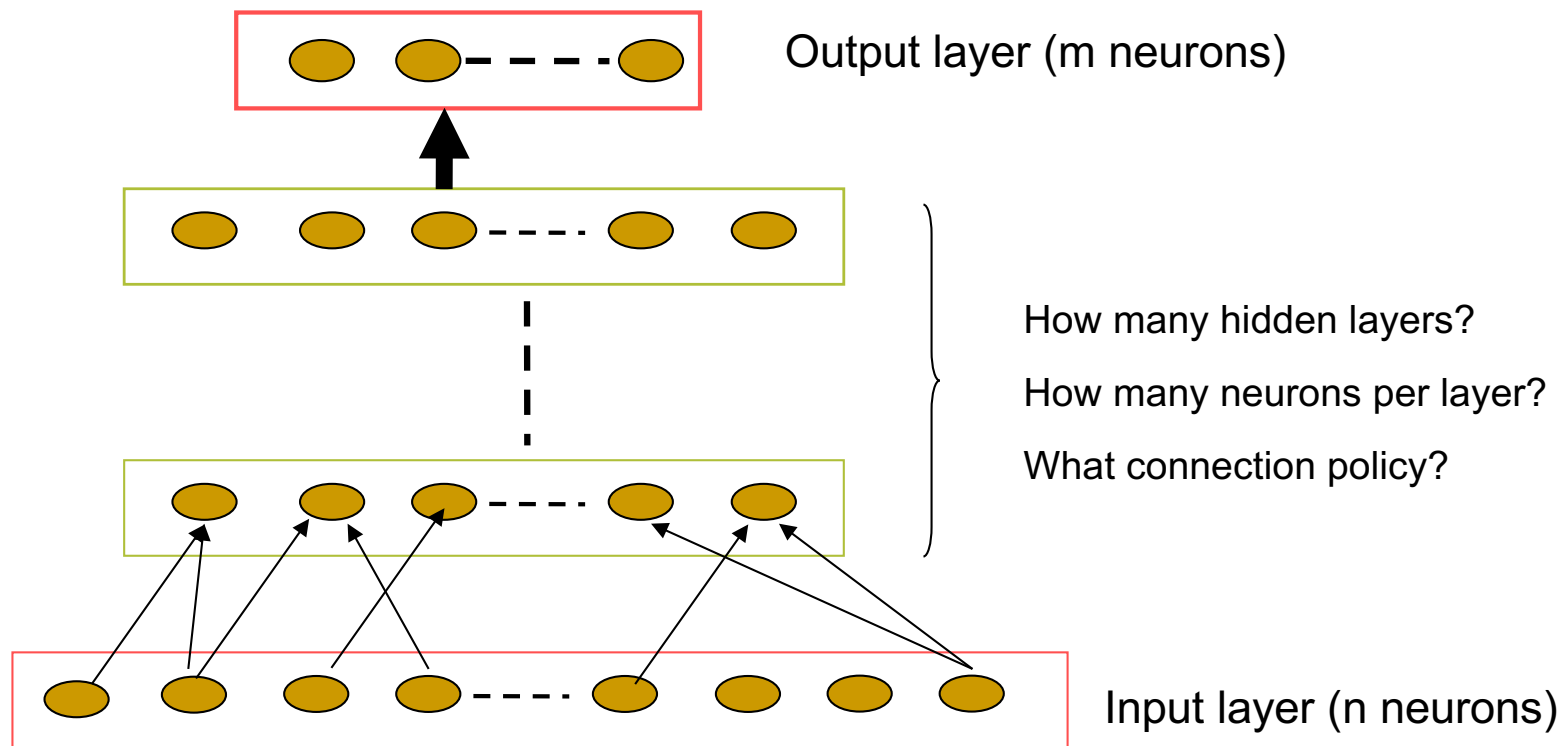
Multilayer perceptron

process



Problem : Settings

Multilayer perceptron *topology*



Problem : Background

- Assumptions :

- MLP Feedforward networks are universal approximators

- K. Hornik, M. Stinchcombe, and H. White, Neural Networks, vol. 2, pp 359-366, 1989

-

Problem : Background

- Existing approaches

- Adhoc approach

- One hidden layer: number of units equal to the average between the number of output units and the number of input units
 -

- Automatic approaches

- Dynamic construction of ANN from the training set
 - Use of an apriori domain knowledge (set of rules)
 - Concept-lattice based ANN
-

Problem : our contribution

Propose an automatic approach of defining interpretable ANN architecture when the domain knowledge is not available

- E. Mephu Nguifo et al., M-CLANN: Multiclass Concept Lattice-Based Artificial Neural Network. **Constructive Neural Networks** 2009: 103-121.
 - Lauraine Tiogning Kueti et al., Boolean factors based Artificial Neural Network. **IJCNN 2016**: 819-825
 - Norbert Tsopzé et al., Towards a generalization of compositional approach of rule extraction from multilayer artificial neural network. **IJCNN 2011**: 1562-1569
-

Our proposal

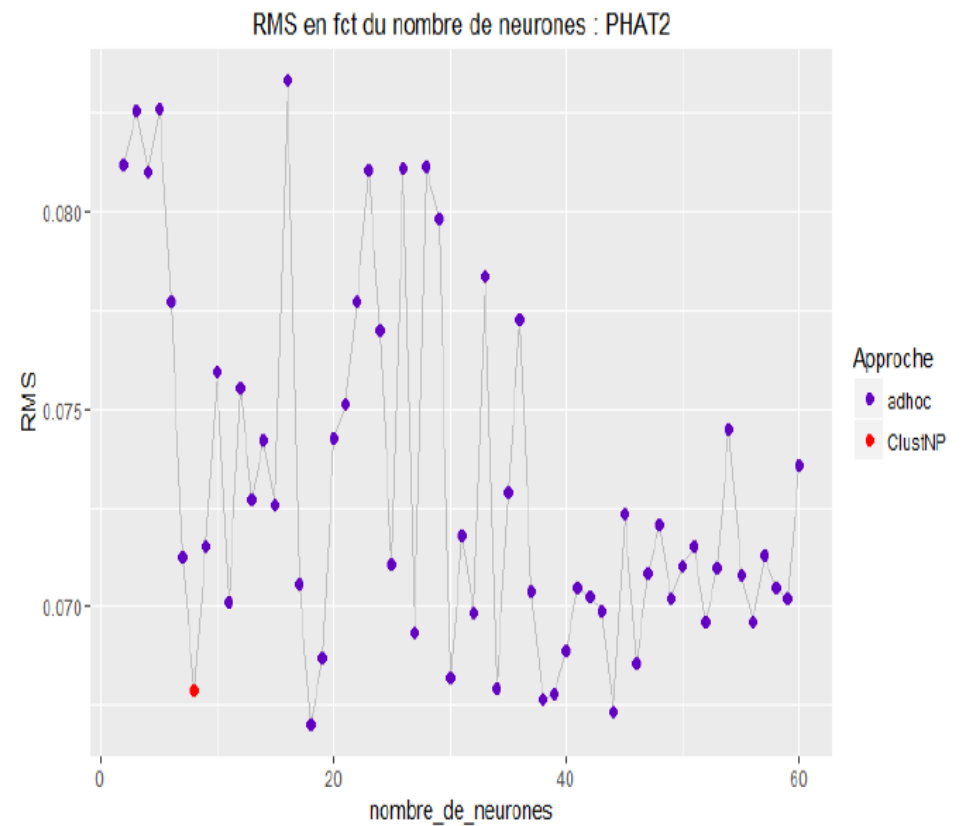
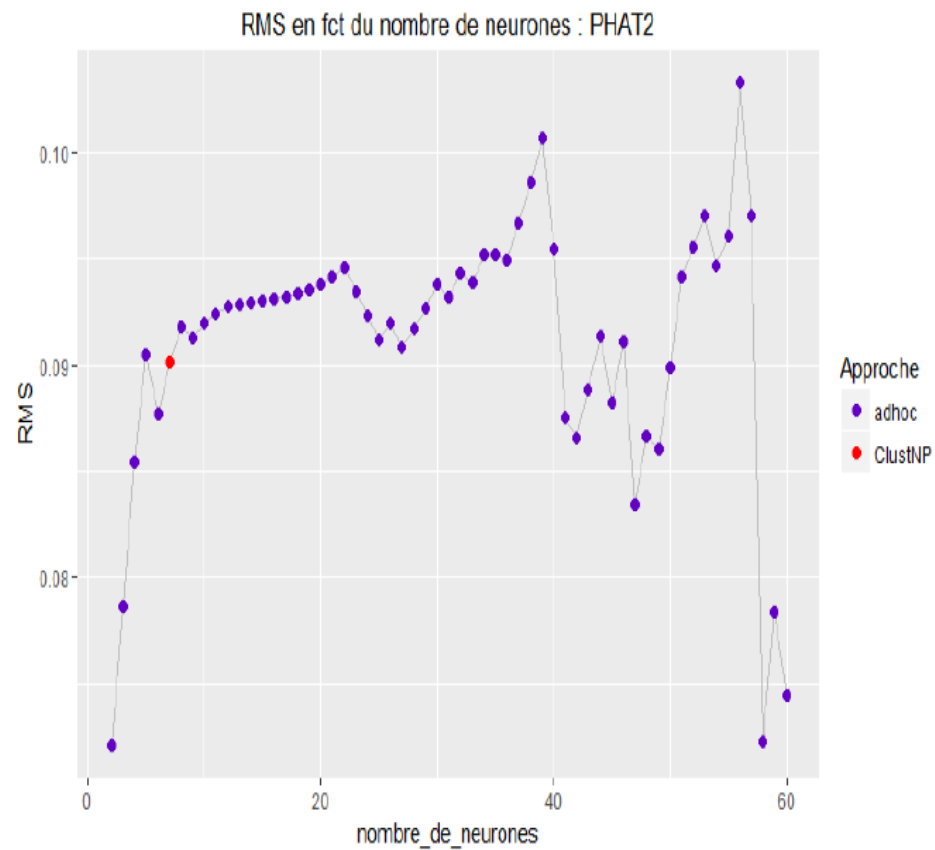
Architecture of ANN :

- Input layer = input data
 - One (neuron) unit for each attribute [+ bias]
 - Hidden layer : one
 - One neuron = one Boolean factor
 - Input layer are fully connected to hidden layer
 - Output = one neuron (Regression)
-

Experimentations

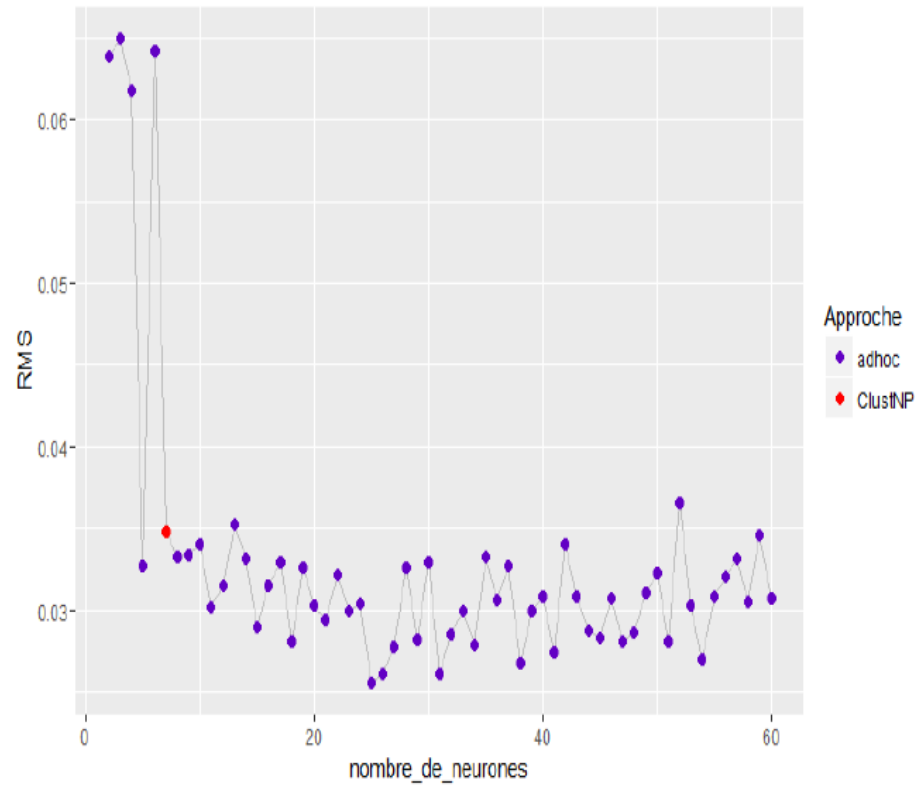
Jeu de données	Nombre d'objets	Taille (après prétraitement)	Nombre de bandes
PHAT2	316	52 Ko	18
PHAT1	515	84 Ko	18
PHAT0	1984	328 Ko	18
SDSS DR9 Data 1	5000	621 Ko	10
SDSS DR9 Data 2	12000	1477 Ko	10
Deep2 DR4 _sans_1	10838	735 Ko	6
Deep2 DR4 _sans_0_1	11784	677 Ko	6
SDSS DR10 Data 1	300 000	32 119 Ko	10
SDSS DR10 Data 2	500 000	53 536 Ko	10

Experimentations

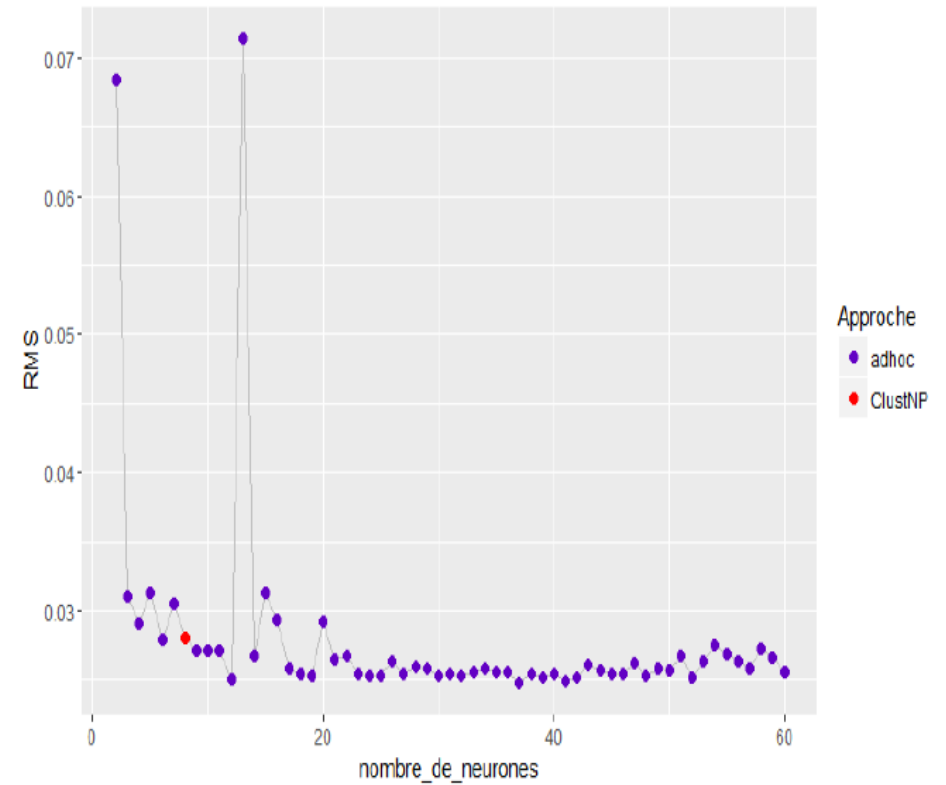


Experimentations

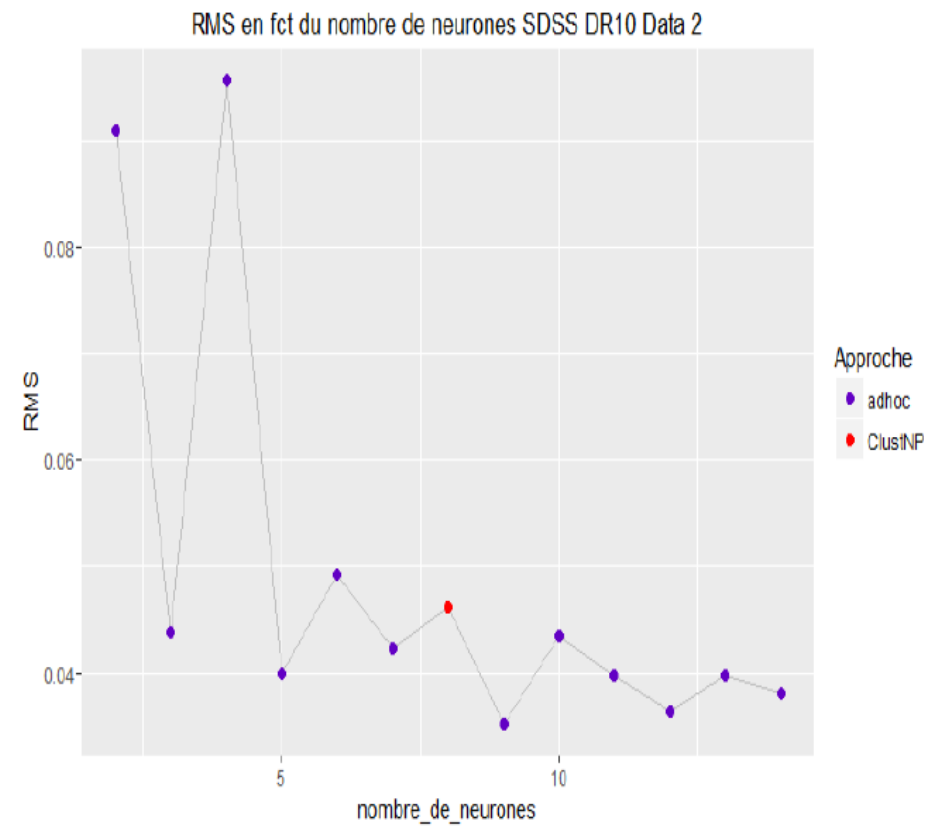
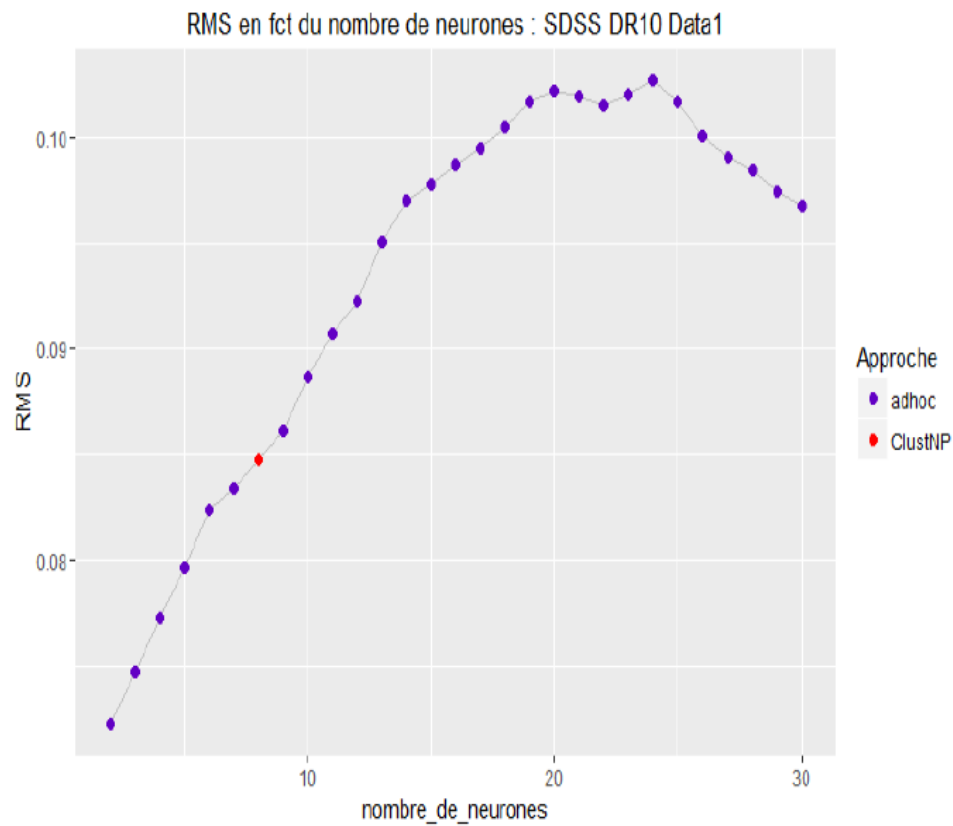
RMS en fct du nombre de neurones : SDSS DR9 Data1



RMS en fct du nombre de neurones : PHAT2



Experimentations (Results)



Experimentations (Results in the Literature)

	class	$\text{std}(\Delta z_{\text{norm}})$	$\text{bias}(\Delta z_{\text{norm}})$	$ \Delta z_{\text{norm}} > 0.15$
[2]	galaxies	0.041	-0.003	0.99%
[3]	galaxies	$\sigma_{68} = 0.03$	-0.001	1.56%
[9]	galaxies	$\sigma_{68} = 0.0248$	0.0008	0.73%
[13]	quasars	0.15	0.032	$> 0.3 : 6.53\%$
[7]	galaxies	0.0490	-0.0081	7.6%
[10]	galaxies	0.024	0.0	1.51%
[6]	galaxies	0.0205	0.00005	4.11%

Conclusion

- BF-ANN, new approach to find **interpretable ANN** architecture when domain knowledge is not available
 - Semantic of neuron
 - Two variants
 - Preliminary validation seems promising
 - Next :
 - Rules extraction (Tsopze et al. IJCNN 2011)
-

Thanks !

