



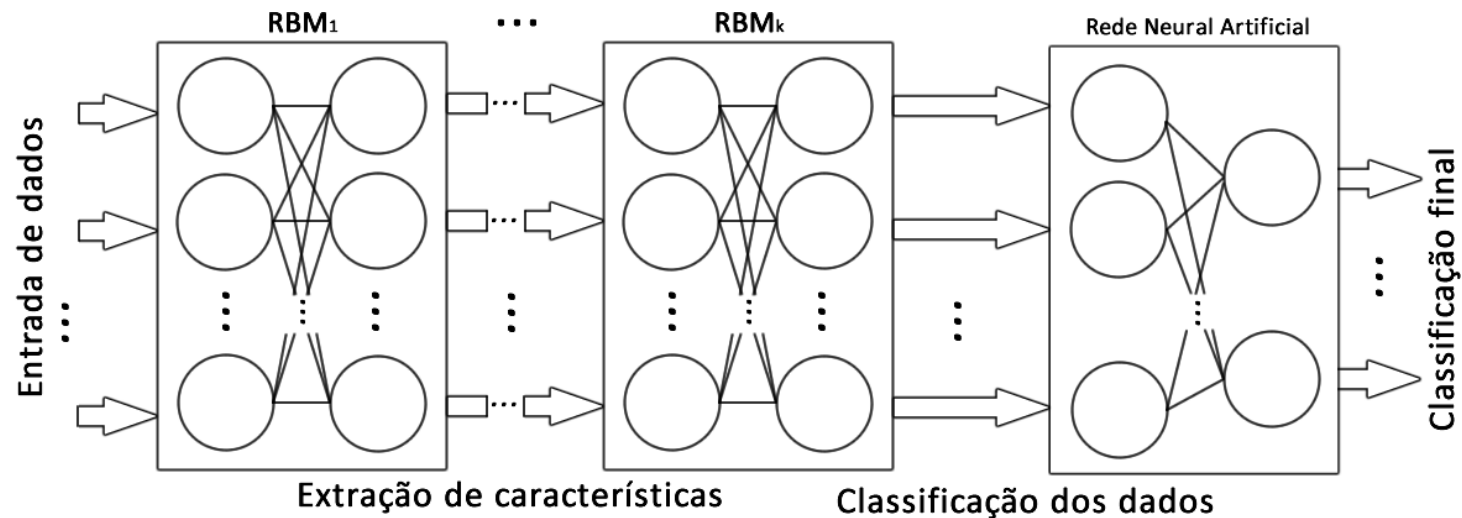
Classificação de grandes bases de dados utilizando máquina de Boltzmann restrita discriminativa

André Pacheco e Renato A. Krohling

Universidade Federal do Espírito Santo – UFES
Programa de Pós-Graduação em Informática – PPGI

- Introdução
- Objetivo
- Máquina de Boltzmann restrita discriminativa
- Resultados experimentais
- Conclusão

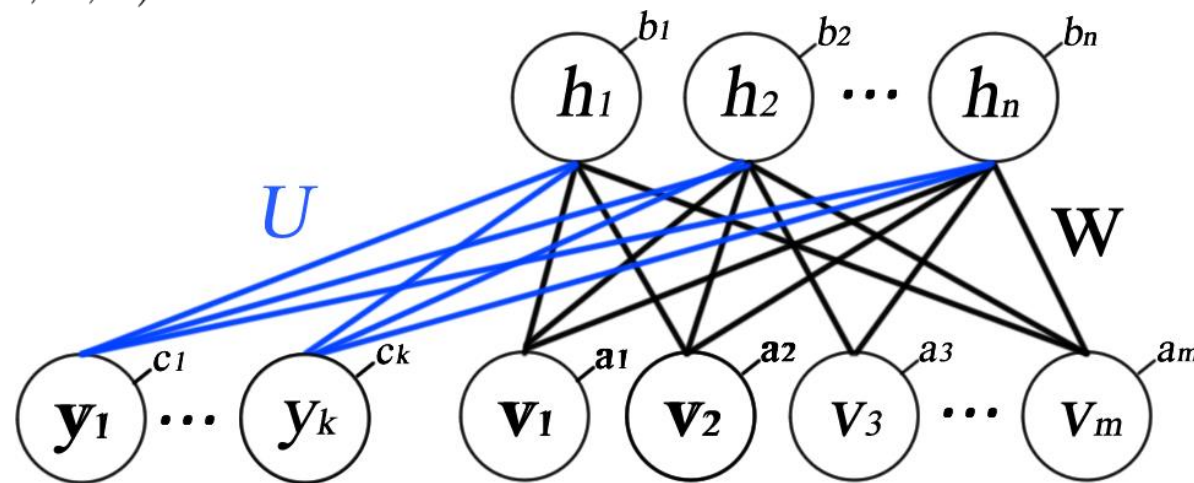
- O problema de classificação de dados é um dos tópicos mais ativos em *machine learning*
 - Está presente em diversos problemas de diversas áreas
 - É um problema supervisionado
 - Muitos problemas possuem padrões de dados difíceis de serem identificados
- Abordagens não supervisionadas, como máquina de Boltzmann restrita (RBM) e autoencoders, vem sendo aplicadas como ferramentas de extração de características a fim de alimentar algoritmos supervisionados



- Baseado na arquitetura híbrida, Larochelle et al. (2012) propuseram a máquina de Boltzmann restrita discriminativa (DRBM)
 - Metodologia baseada na RBM
 - Abordagem semi-supervisionada: engloba as abordagens não supervisionada e supervisionada
- A DRBM é capaz de extrair padrões do dados, assim com a RBM, e classificar as amostras sem necessidade de acoplar outro algoritmo
- Vem sendo utilizada para classificação de dados extraídos de imagens, principalmente classificação de dígitos escritos à mão

- O principal objetivo deste trabalho é aplicar a DRBM para grandes bases de dados que não sejam extraídas de imagens
 - As grandes bases possuem dois tipos: muitas amostras ou muitas *features*
- Compreender qual gama de problemas é interessante aplicar a DRBM

- A DRBM é basicamente uma rede estocástica constituída por três camadas: visível (v), oculta (h) e a de rótulos (y). Em relação a conexão entre os neurônios
 - São bidirecionais e simétricos
 - Não existe conexão entre neurônios de mesma camada
- Os pesos de ligação entre as camadas e os vetores *bias* são organizado como $(\mathbf{W}, \mathbf{U}, \mathbf{a}, \mathbf{b}, \mathbf{c})$ denominado θ



- A probabilidade conjunta da configuração $(\mathbf{y}, \mathbf{v}, \mathbf{h})$ é descrita por:

$$p(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = \frac{e^{-E(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta})}}{\sum_{\mathbf{y}, \mathbf{v}, \mathbf{h}} e^{-E(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta})}}$$

- Na qual a função de energia é dada por:

$$E(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = - \sum_{i=1}^m \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^n b_j h_j - \sum_{z=1}^k c_z y_z - \sum_{i,j=1}^{m,n} \frac{v_i}{\sigma^2} h_j w_{ij} - \sum_{z,j=1}^{k,n} y_z h_j u_{zj}$$

- A partir da distribuição conjunta são calculadas as distribuições condicionais para cada camada:

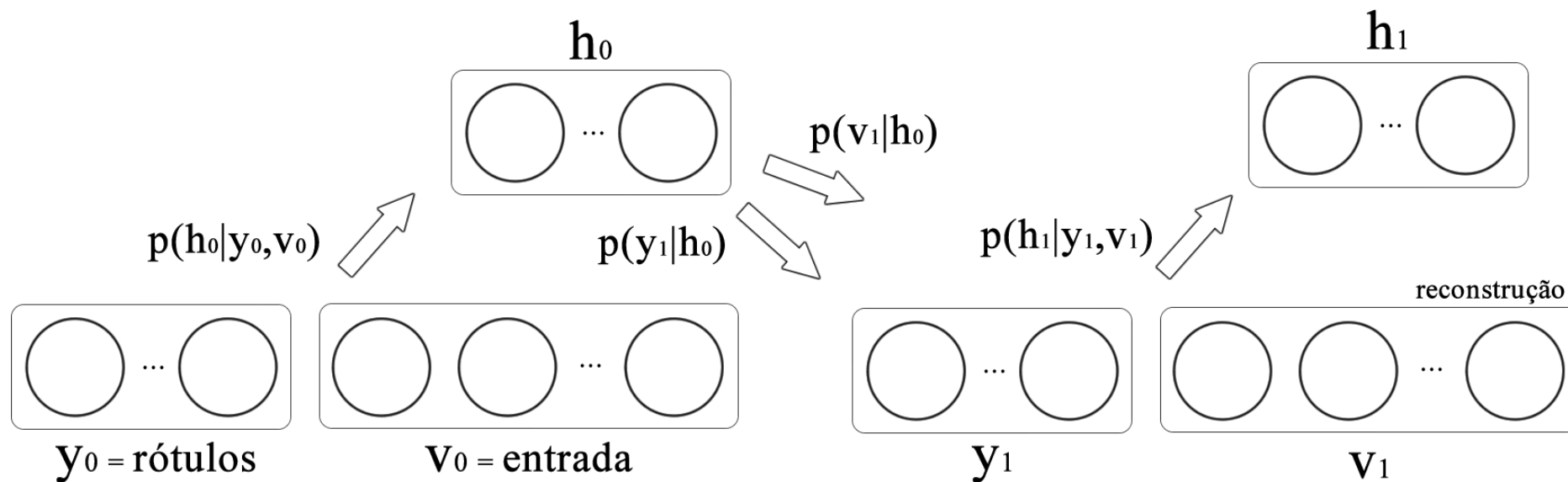
$$p(v_i = v | \mathbf{h}; \boldsymbol{\theta}) = N(v | a_i + \sum_{j=1}^n h_j w_{ij}, \sigma^2)$$

$$p(h_j = 1 | \mathbf{y}, \mathbf{v}; \boldsymbol{\theta}) = \phi(b_j + \sum_{z=1}^k y_z u_{zj} + \sum_{i=1}^m v_i w_{ij})$$

$$p(y_z = 1 | \mathbf{h}; \boldsymbol{\theta}) = \frac{e^{(c_z + \sum_{j=1}^n h_j u_{zj})}}{\sum_{l=1}^k e^{(c_l + \sum_{j=1}^n h_j u_{lj})}}$$

- O treinamento da DRBM visa ajustar os parâmetros $(\mathbf{W}, \mathbf{U}, \mathbf{a}, \mathbf{b}, \mathbf{c})$ de modo a minimizar a energia do sistema

- O algoritmo de treinamento utilizado é a divergência contrastiva (CD):



- Baseado nos passos do algoritimo CD, as regras de atualização para os $p(\mathbf{W}, \mathbf{U}, \mathbf{a}, \mathbf{b}, \mathbf{c})$ são descritas por:

Pesos de conexão

$$\mathbf{W}^{t+1} = \mathbf{W}^t + \Delta\mathbf{W}^t \rightarrow \Delta\mathbf{W}^t = \hat{\eta}(\mathbf{v}_0\mathbf{h}_0^T - \mathbf{v}_1\mathbf{h}_1^T) - \hat{\lambda}\mathbf{W}^t + \hat{\alpha}\Delta\mathbf{W}^{t-1}$$

$$\mathbf{U}^{t+1} = \mathbf{U}^t + \Delta\mathbf{U}^t \rightarrow \Delta\mathbf{U}^t = \hat{\eta}(\mathbf{y}_0\mathbf{h}_0^T - \mathbf{y}_1\mathbf{h}_1^T) - \hat{\lambda}\mathbf{U}^t + \hat{\alpha}\Delta\mathbf{U}^{t-1}$$

Bias

$$\mathbf{a}^{t+1} = \mathbf{a}^t + \Delta\mathbf{a}^t \rightarrow \Delta\mathbf{a}^t = \hat{\eta}(\mathbf{v}_0 - \mathbf{v}_1) + \hat{\alpha}\Delta\mathbf{a}^{t-1}$$

$$\mathbf{b}^{t+1} = \mathbf{b}^t + \Delta\mathbf{b}^t \rightarrow \Delta\mathbf{b}^t = \hat{\eta}(\mathbf{h}_0 - \mathbf{h}_1) + \hat{\alpha}\Delta\mathbf{b}^{t-1}$$

$$\mathbf{c}^{t+1} = \mathbf{c}^t + \Delta\mathbf{c}^t \rightarrow \Delta\mathbf{c}^t = \hat{\eta}(\mathbf{y}_0 - \mathbf{y}_1) + \hat{\alpha}\Delta\mathbf{c}^{t-1}$$

onde $\hat{\eta}$, $\hat{\lambda}$ e $\hat{\alpha}$ são constantes conhecidas como taxa de aprendizado, fator de decaimento e *momentum*

- A classificação final de uma entrada \mathbf{x} será a classe com maior probabilidade retornada pela equação:

$$p(y^e | \mathbf{x}) = \frac{e^{-F(\mathbf{x}, y^e)}}{\sum_{y^* \in \{1, 2, \dots, k\}} e^{-F(\mathbf{x}, y^*)}}$$

onde $F(\mathbf{x}, y^e)$ é conhecida como energia livre e pode ser calculada como:

$$F(\mathbf{x}, \mathbf{y}) = c_e + \sum_{j=1}^n \varphi\left(\sum_{i=1}^m x_i w_{ij} + u_{yj} + b_j\right)$$

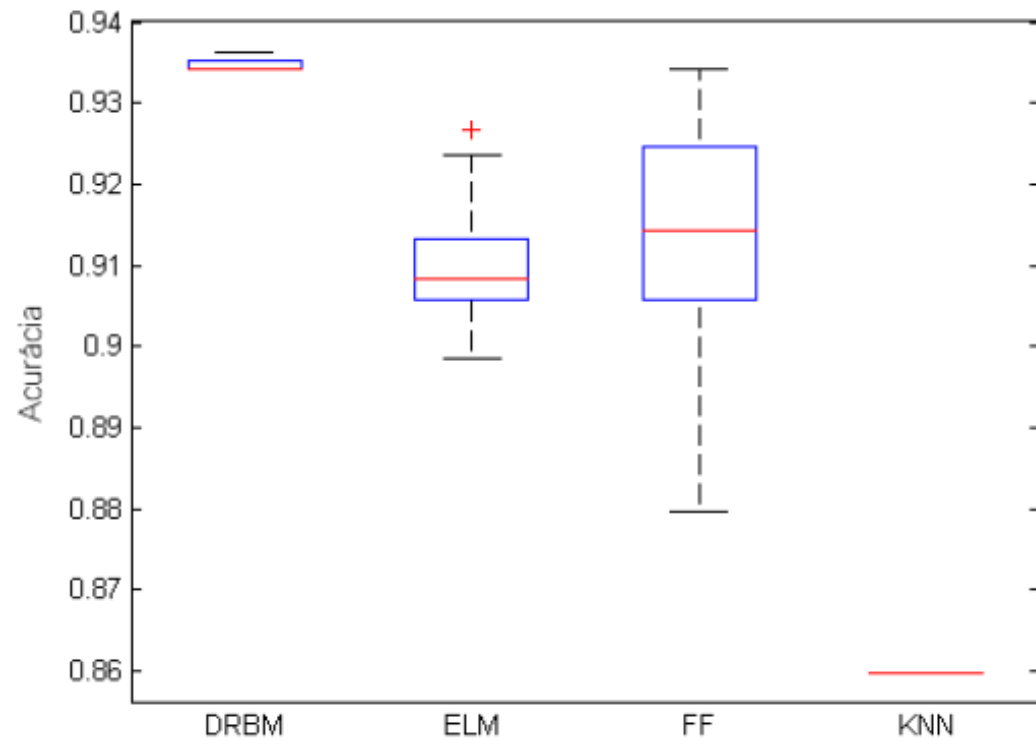
- Os experimentos foram realizados utilizando 5 bases de dados do repositório UCI

Base de Dados	# de amostras	# de atributos	# de classes
<i>DNA</i>	3186	180	3
<i>Covtype</i>	581012	54	7
<i>Higgs</i>	1000000	28	2
<i>Isolet</i>	7797	617	26
<i>Susy</i>	1000000	18	2

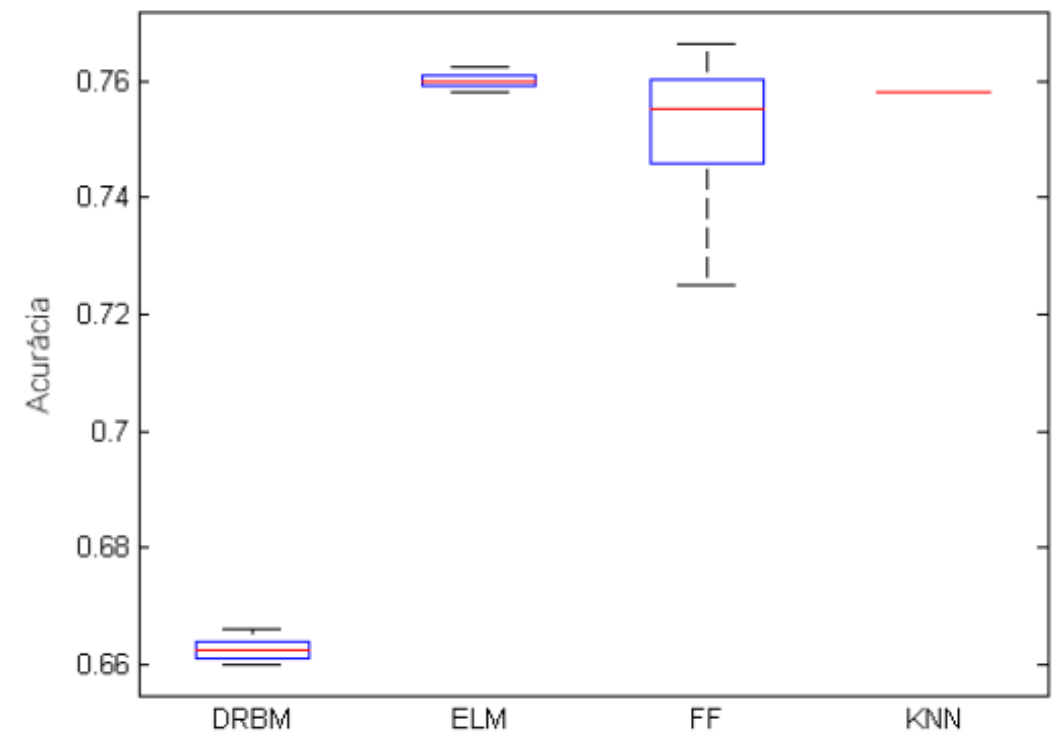
- A DRBM utilizada possui 500 neurônios na camada oculta, utiliza mini-batch de 100 partições, realiza no máximo 300 iterações.
- Os resultados da DRBM foram comparados com os seguintes algoritmos:
 - Uma rede neural *feedforward* com treinamento via Levenberg-Marquardt *backpropagation*
 - Uma máquina de aprendizado extremo (ELM)
 - Um K-vizinhos mais próximos (KNN)

- As bases de dados foram divididas em 70% para treinamento e 30% para testes
- Cada algoritmo, com exceção do KNN, foi executados 30 vezes para cálculo estatístico
 - Desempenho em termos de média e desvio padrão da acurácia de classificação

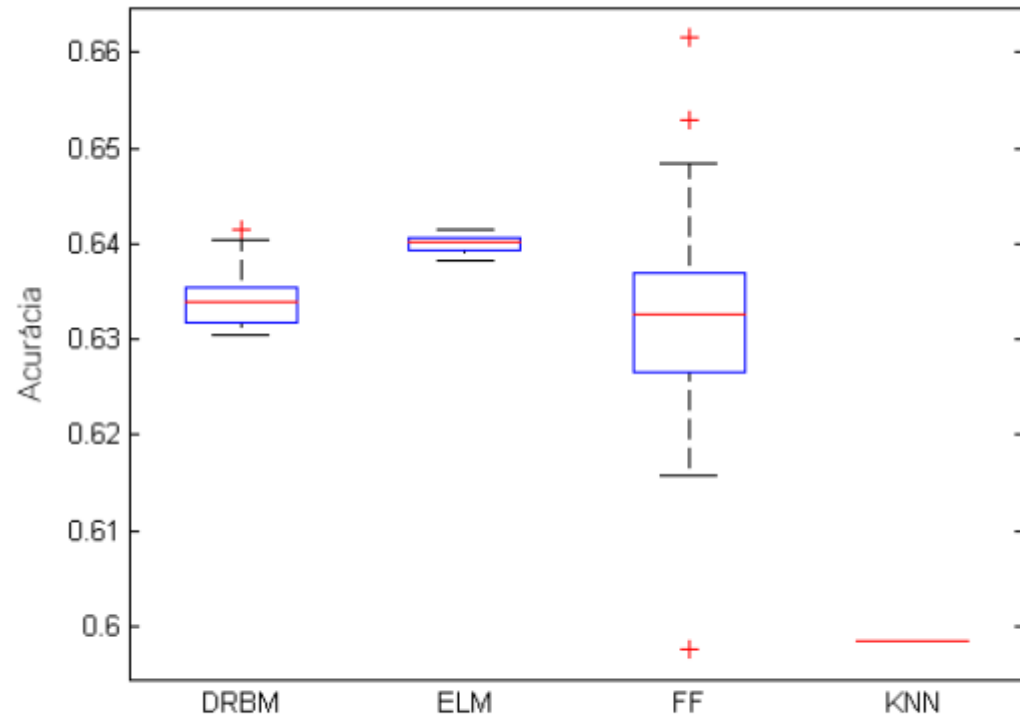
Acurácia em %				
Base de Dados	DRBM	ELM	FF	KNN
<i>DNA</i>	93,45 ± 0,07	90,39 ± 0,75	91,36 ± 1,34	85,98 ± 0
<i>Covtype</i>	66,25 ± 0,17	76,01 ± 0,11	75,22 ± 1,09	75,81 ± 0
<i>Higgs</i>	63,40 ± 0,30	63,99 ± 0,09	63,21 ± 1,19	59,84 ± 0
<i>Isolet</i>	93,74 ± 0,16	86,81 ± 0,60	89,41 ± 1,70	88,24 ± 0
<i>Susy</i>	76,39 ± 0,32	79,39 ± 0,29	78,14 ± 0,65	70,88 ± 0



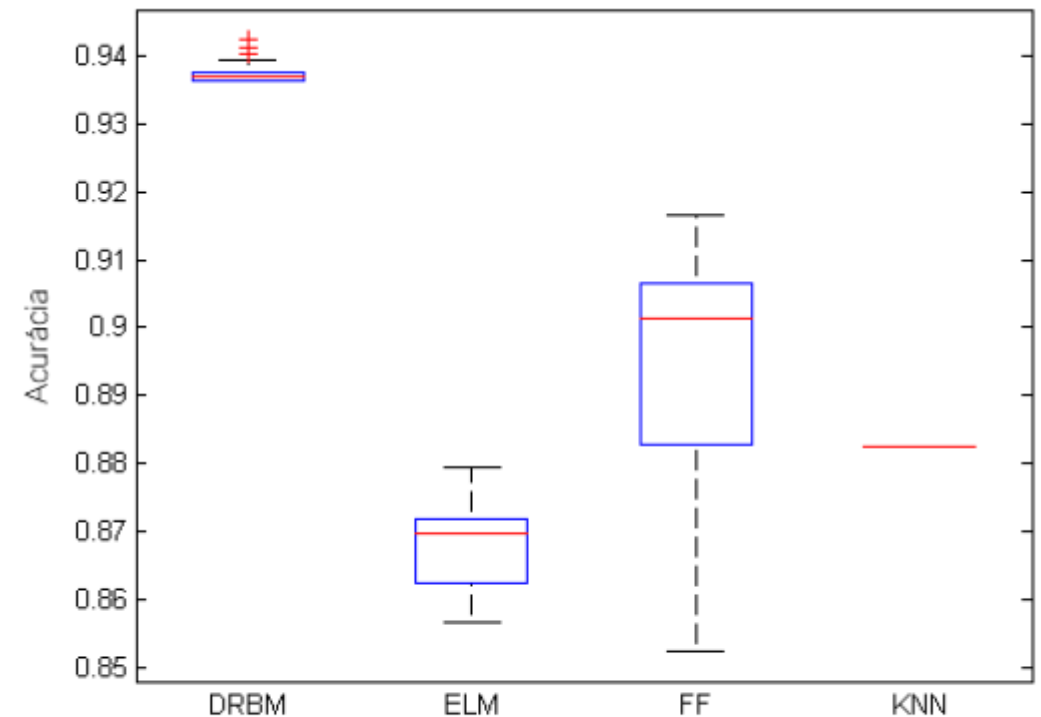
(a) *DNA*



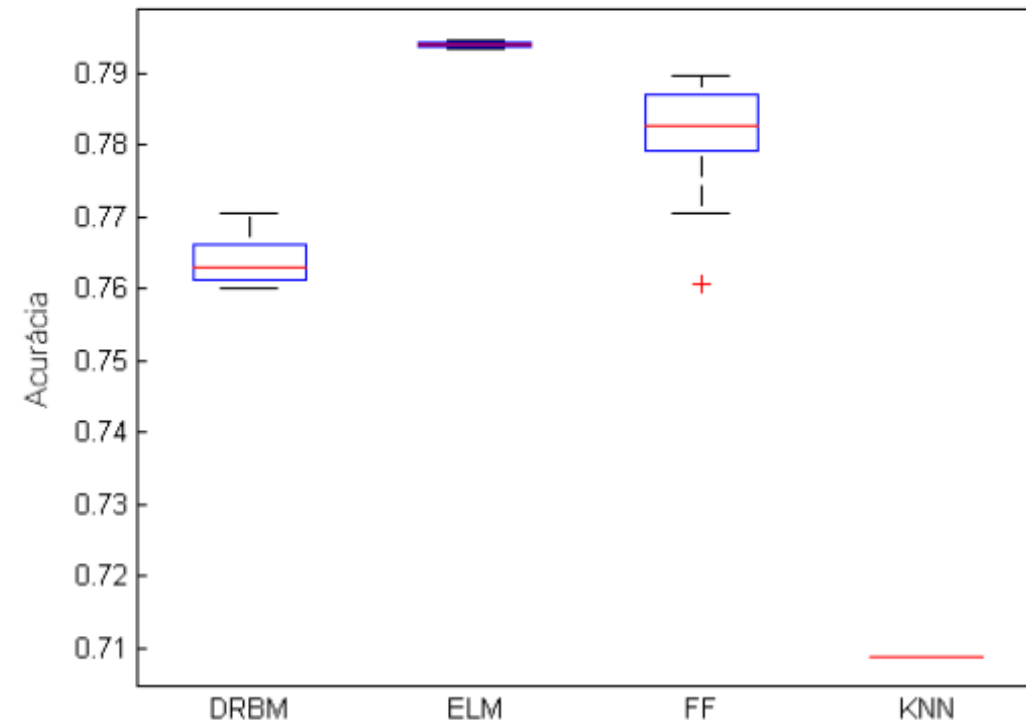
(b) *Covtype*



(c) *Higgs*



(d) *Isolet*



(e) *Susy*

- A máquina de Boltzmann restrita discriminativa foi aplicada para 5 bases de dados não extraídas de imagens
 - Os resultados da DRBM foram comparados com 3 algoritmos estabelecidos na literatura
- De maneira geral, o desempenho da DRBM foi competitivo com os demais
- Todavia, os resultados apontaram que a DRBM é uma boa escolha quando a base dados possui um número elevado de *features*
- Esse resultado é coerente com o fato de que a RBM vem sendo utilizada para reduzir dimensionalidade de grandes bases de dados

Hinton, G. E., Osindero, S., e Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7): 1527–1554.

Larochelle, H. e Bengio, Y. (2008). Classification using discriminative restricted Boltzmann machines. In *Proceedings of the 25th international conference on Machine learning*, p. 536–543. ACM.

Larochelle, H., Mandel, M., Pascanu, R., e Bengio, Y. (2012). Learning algorithms for the classification restricted Boltzmann machine. *The Journal of Machine Learning Research*, 13(1): 643–669.

Papa, J. P., Rosa, G. H., Marana, A. N., Scheirer, W., e Cox, D. D. (2015). Model selection for discriminative restricted Boltzmann machines through meta-heuristic techniques. *Journal of Computational Science*, 9:14–18.

Tamilselvan, P. e Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, 115:124–135.



Obrigado pela atenção

Contato:

pacheco.comp@gmail.com

Krohling.renato@gmail.com