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

André G. C. Pacheco
UFES - Universidade Federal do Espírito Santo
PPGI - Programa de Pós-Graduação em Informática
Av. Fernando Ferrari 514 - CEP 29060-270, Vitória, ES, Brasil
pacheco.comp@gmail.com

Renato A. Krohling
UFES - Universidade Federal do Espírito Santo
PPGI - Programa de Pós-Graduação em Informática
Av. Fernando Ferrari 514 - CEP 29060-270, Vitória, ES, Brasil
krohling.renato@gmail.com

RESUMO

Máquinas de Boltzmann restrita vem sendo aplicadas em diversos problemas de aprendizagem. Na maior parte deles, o método é utilizado como etapa de extração de características com objetivo de alimentar outro algoritmo supervisionado. Baseado nisso, foi desenvolvido recentemente a máquina de Boltzmann restrita discriminativa, no qual a ideia principal é unir o aprendizado não supervisionado e o supervisionado em um mesmo método. Desde então, essa nova abordagem vem sendo utilizada para classificar, principalmente, bases de dados extraídas de imagens. Sendo assim, neste trabalho a máquina de Boltzmann restrita discriminativa é aplicada para classificar grandes bases de dados não advindas de imagens e seus resultados se mostraram competitivos com outros classificadores clássicos da literatura.

PALAVRAS CHAVE. Máquina de Boltzmann restrita. Máquina de Boltzmann restrita discriminativa. Classificação de dados.

Tópicos: Modelos probabilísticos e Estatística

ABSTRACT

Restricted Boltzmann machines have been applied for a variety of learning problems. For most of them, the method is used as a feature extraction step in order to feed another supervised algorithm. Discriminative restricted Boltzmann machine was recently developed, which the main idea is to join the unsupervised learning and the supervised learning in the same method. Since then, this new approach has been used to classify, mainly, data bases extracted from images. Thus, in this work the discriminative restricted Boltzmann machines is applied to classify large data bases, which are not extracted from images, and its results are competitive with others standard classifiers.

KEYWORDS. Restricted Boltzmann machine. Discriminative restricted Boltzmann machine. Data classification.

Paper topics: Probabilistic models and Statistics

1. Introdução

O problema de classificação de dados está intrinsecamente relacionado com o reconhecimento de padrões e regularidades de uma determinada base de dados. No contexto de sistemas de aprendizagem, classificar dados é considerado um problema supervisionado. Porém, abordagens não supervisionadas, como máquina de Boltzmann restrita [Hinton, 2002; Smolensky, 1986] e autoencoders [Bourlard e Kamp, 1988], vem sendo aplicadas como ferramentas de extração de característica a fim de alimentar algoritmos supervisionados, como redes neurais artificiais [Haykin, 2007]. Desta forma, surgem as técnicas semi-supervisionadas, que ganharam destaques nos últimos anos.

A máquina de Boltzmann restrita (do inglês: *restricted Boltzmann machine*, RBM) é uma rede estocástica amplamente utilizada para compor redes de crença profundas (do inglês: *deep belief networks*, DBN) [Hinton et al., 2006]. A RBM é capaz de extrair características de um conjunto de dados por meio de treinamento não supervisionado. Devido a isso, abordagens que utilizam RBM's, para compor uma DBN, foram desenvolvidas como primeiro estágio de um classificador baseado em redes neurais artificiais [Salama et al., 2010; Tamilselvan e Wang, 2013]. A Figura 1 mostra uma arquitetura híbrida baseada em composições de RBM's alimentando uma rede neural, constituindo assim, uma abordagem semi-supervisionada.

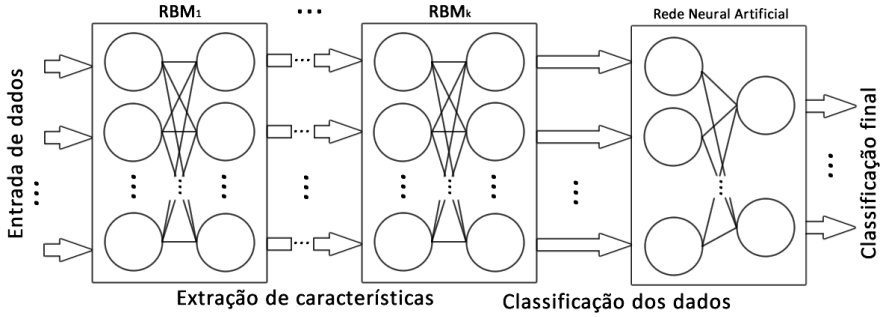


Figura 1: Arquitetura híbrida semi-supervisionada

Baseado na arquitetura híbrida, Larochelle e Bengio [2008] e Larochelle et al. [2012] propuseram um modelo completo, baseado em RBM, que engloba tanto a abordagem supervisionada, quanto a não supervisionada, denominado máquina de Boltzmann restrita discriminativa (do inglês: *discriminative restricted Boltzmann machine*, DRBM). Além dos dados, a DRBM necessita dos rótulos de classificação de cada amostra, com isso o algoritmo é capaz de extrair as características do conjunto, assim como uma RBM, e classificar a amostra, sem a necessidade de acoplar um outro algoritmo supervisionado à abordagem.

Atualmente, um dos grandes desafios da computação é desenvolver algoritmos capazes de processar e entender características de grandes bases de dados. Até o momento a DRBM vem sendo utilizada para classificação de dados extraídos de imagens, principalmente classificação de dígitos escritos à mão [Larochelle et al., 2012; Ji et al., 2014; Papa et al., 2015]. O objetivo principal deste trabalho é verificar a eficácia do método para grandes bases de dados que não sejam extraídas de imagens. O resultados obtidos pela DRBM foram comparados com três algoritmos amplamente utilizados na área, uma rede ELM (do inglês: *extreme learning machine*) [Huang et al., 2006], uma rede neural *feedforward* [Haykin, 2007] e um KNN (do inglês: *k-nearest neighbors*) [Keller et al., 1985] e mostraram que a DRBM é competitiva com os mesmos. O restante deste artigo está organizado da seguinte forma: na seção 2 são apresentados os conceitos básicos referentes a RBM. Na seção 3 a DRBM é descrita. Por fim, nas seções 3 e 4, são apresentados os resultados experimentais e uma breve conclusão, respectivamente.

2. Máquina de Boltzmann restrita

2.1. Conceitos Básicos

A máquina de Boltzmann restrita [Smolensky, 1986; Hinton, 2002] é, basicamente, uma rede estocástica constituída por duas camadas: visível e oculta. A camada de unidades visíveis representam os dados observados e está conectada à camada oculta, que por sua vez, deverá aprender a extrair características desses dados [Memisevic e Hinton, 2010]. Originalmente, a RBM foi desenvolvida para dados binários, tanto na camada visível quanto na camada oculta. Essa abordagem é conhecida como Bernoulli-Bernoulli RBM (BBRBM). Devido ao fato de existir problemas onde é necessário processar outros tipos de dados, Hinton e Salakhutdinov [2006] propuseram a Gaussian-Bernoulli RBM (GBRBM), que utiliza uma distribuição normal para modelar os neurônios da camada visível. Nesta seção, serão descritos os conceitos básicos referentes a abordagem GBRBM.

Na RBM as conexões entre neurônios são bidirecionais e simétricas. Isso significa que existe tráfego de informação em ambos os sentidos da rede. Além disso, para simplificar procedimentos de inferência, neurônios de uma mesma camada não estão conectados entre si. Sendo assim, só existe conexão entre neurônios de camadas diferentes, por isso a máquina é restrita. Na Figura 2 é mostrado uma RBM com m neurônios na camada visível $(v_1, ..., v_m)$, n neurônios na camada oculta $(h_1, ..., h_n)$, sendo $(a_1, ..., a_m)$ e $(b_1, ..., b_n)$ os vetores de *bias* e por fim $\mathbf{W}$ a matriz de pesos das conexões. Daqui até o fim da seção 2, o conjunto $(\mathbf{W}, \mathbf{a}, \mathbf{b})$ será denominado $\boldsymbol{\theta}$.

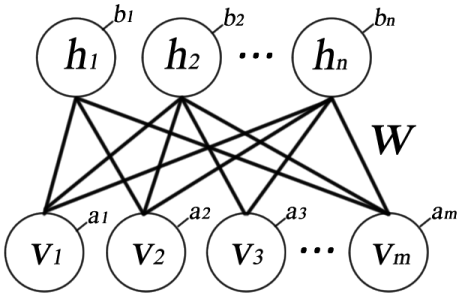


Figura 2: A máquina de Boltzmann restrita

A RBM é um modelo baseado em energia, sendo que a distribuição de probabilidade conjunta da configuração (**v**,**h**) é descrita por:

$$p(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = \frac{e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}}{\sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\boldsymbol{\theta})}} \tag{1}$$

onde a função de energia descrita por:

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

A probabilidade que a rede atribui a um vetor visível, $\mathbf{v}$, é dada pela soma de todas as probabilidades dos vetores escondidos $\mathbf{h}$, calculados por:

$$p(\mathbf{v}; \boldsymbol{\theta}) = \frac{\sum_{\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\theta)}}{\sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\theta)}} \tag{3}$$

Como a RBM é restrita, ou seja, não possui conexões de neurônios entre uma mesma camada, as distribuições de probabilidade de $\mathbf{h}$ dado $\mathbf{v}$ e de $\mathbf{v}$ dado $\mathbf{h}$ são descritas pelas equações 4 e 5, respectivamente.

$$p(\mathbf{h}|\mathbf{v};\boldsymbol{\theta}) = \prod_j p(h_j|\mathbf{v}) \tag{4}$$

$$p(\mathbf{v}|\mathbf{h};\boldsymbol{\theta}) = \prod_i p(v_i|\mathbf{h}) \tag{5}$$

Baseado na versão GBRBM [Hinton e Salakhutdinov, 2006], no qual a camada visível é contínua e a escondida binária, as distribuições condicionais são descritas pelas equações 6 e 7.

$$p(h_j = 1|\mathbf{v};\boldsymbol{\theta}) = \phi(b_j + \sum_{i=1}^{m} v_i w_{ij}) \tag{6}$$

$$p(v_i = v|\mathbf{h};\boldsymbol{\theta}) = N(v|a_i + \sum_{j=1}^{n} h_j w_{ij}, \sigma^2) \tag{7}$$

onde $\phi(x) = \frac{1}{1+e^{-x}}$ é a função logística, e N é uma distribuição normal, com média v e desvio padrão σ^2, normalmente utilizado como 1.

2.2. Treinando uma RBM

Basicamente, o objetivo do treinamento da RBM é estimar $\boldsymbol{\theta}$ que faça com que a energia da rede diminua [Hinton, 2012]. Como $p(\mathbf{v};\boldsymbol{\theta})$ é a distribuição dos dados de entrada, $\boldsymbol{\theta}$ pode ser estimado a partir da maximização de $p(\mathbf{v},\boldsymbol{\theta})$ ou, de maneira equivalente, log $p(\mathbf{v},\boldsymbol{\theta})$. Sendo assim, o gradiente descendente de log $p(\mathbf{v},\boldsymbol{\theta})$, com respeito a $\boldsymbol{\theta}$ é calculado da seguinte maneira:

$$\frac{\partial p(\mathbf{v},\boldsymbol{\theta})}{\partial \theta} = \langle v_i h_j \rangle_d - \langle v_i h_j \rangle_m \tag{8}$$

onde as componentes $\langle v_i h_j \rangle_d$ e $\langle v_i h_j \rangle_m$ são usadas para denotar as expectativas computadas sob os dados e o modelo, respectivamente.

Obter uma estimativa de $\langle v_i h_j \rangle_d$ é realizada de maneira simples através das equações 6 e 7. Todavia, obter uma estimativa de $\langle v_i h_j \rangle_m$ é muito mais difícil. Isso pode ser feito por meio do amostrador de Gibbs utilizando dados aleatórios alimentando a camada visível. Todavia, esse procedimento pode consumir um longo tempo para obter um resultado adequado. Felizmente, um procedimento mais rápido, denominado *contrastive divergence* (CD), foi proposto por Hinton [2002] cuja ideia é alimentar a camada visível com dados de treinamento e executar o amostrador de Gibbs apenas uma vez, como ilustrado na Figura 3. Hinton [2002] denominou esta etapa de reconstrução. O algoritmo CD é descrito na forma de passos a seguir:

Passo 1: Igualar a camada visível $\mathbf{v_0}$ aos dados de entradas e, logo após, estimar a camada escondida $\mathbf{h_0}$ por meio da equação 6. Com isso, $\left\langle \mathbf{v}\mathbf{h^T} \right\rangle_d = \mathbf{v_0}\mathbf{h_0^T}$.

Passo 2: A partir de $\mathbf{h_0}$ estimar $\mathbf{v_1}$ por meio da equação 7. Essa etapa é conhecida como reconstrução, pois os dados de entrada serão reconstruídos.

Passo 3: A partir de $\mathbf{v_1}$ estimar $\mathbf{h_1}$, novamente por meio da equação 6. Com isso, $\left\langle \mathbf{v}\mathbf{h^T} \right\rangle_m = \mathbf{v_1}\mathbf{h_1^T}$.

Passo 4: O conjunto de parâmetros $comoilustraaFigura\theta$ são atualizados da seguinte forma:

$$\mathbf{W^{t+1}} = \mathbf{W^t} + \Delta\mathbf{W^t} \rightarrow \Delta\mathbf{W^t} = \eta(\mathbf{v_0}\mathbf{h_0^T} - \mathbf{v_1}\mathbf{h_1^T}) - \lambda\mathbf{W^t} + \alpha\Delta\mathbf{W^{t-1}} \tag{9}$$

$$\mathbf{a^{t+1}} = \mathbf{a^t} + \Delta\mathbf{a^t} \rightarrow \Delta\mathbf{a^t} = \eta(\mathbf{v_0} - \mathbf{v_1}) + \alpha\Delta\mathbf{a^{t-1}} \tag{10}$$

$$\mathbf{b^{t+1}} = \mathbf{b^t} + \Delta\mathbf{b^t} \rightarrow \Delta\mathbf{b^t} = \eta(\mathbf{h_0} - \mathbf{h_1}) + \alpha\Delta\mathbf{b^{t-1}} \tag{11}$$

onde $(\mathbf{W}, \mathbf{a}, \mathbf{b})$ são inicializados de maneira aleatória.

Passo 5: repetir os passos 2, 3 e 4 em um número predeterminado de iterações.

Os parâmetros η, λ e α são conhecidos como taxa de aprendizado, fator de decaimento e *momentum*. Hinton [2002] sugere $\eta = 0.01$, $\lambda = [0.01, 0.0001]$ e $\alpha = 0.5$ para iteração menor do que 5 e $\alpha = 0.9$ caso contrário. Trabalhos relacionados a escolhas destes parâmetros tem sido desenvolvidos [Liu et al., 2014; Papa et al., 2015, 2016].

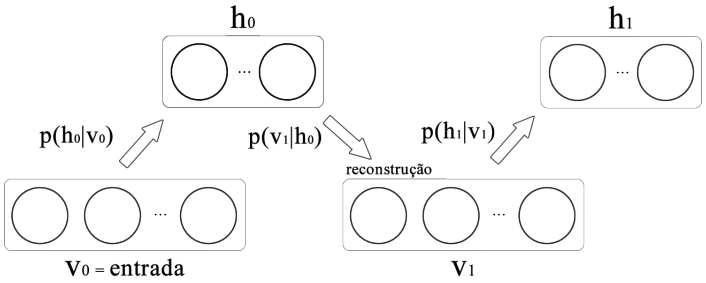


Figura 3: Processo de reconstrução do algoritmo CD para RBM

3. Máquina de Boltzmann restrita discriminativa

3.1. Conceitos básicos

A máquina de Boltzmann restrita discriminativa é, basicamente, um incremento na RBM com intuito de torná-la um algoritmo semi-supervisionado para problemas de classificação. A ideia principal do método é incorporar na RBM os rótulos de cada classe na camada de entrada, e com isso, determinar a distribuição conjunta, tanto dos dados, quanto dos rótulos, de uma maneira discriminativa, ou seja, calcular a probabilidade de cada classe dado uma certa amostra [Papa et al., 2015].

Além das características comuns à RBM, a DRBM possui uma camada de rótulos, representada por $\mathbf{y}$, os pesos de das conexões entre $\mathbf{y}$ e $\mathbf{h}$, representado por $\mathbf{U}$ e o vetor de bias, $\mathbf{c}$. A Figura 4 ilustra uma DRBM no qual as ligações entre a camada de rótulos e a camada oculta estão destacadas em azul. O conjunto $\boldsymbol{\theta}$, definido da seção 2, será acrescido de $\mathbf{c}$ e $\mathbf{U}$.

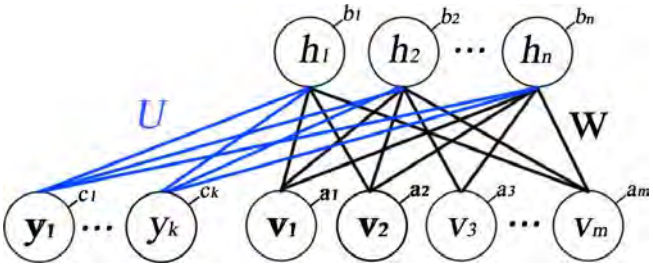


Figura 4: A máquina de Boltzmann restrita discriminativa

O modelo matemático da DRBM é bastante similar ao da RBM. Sendo assim, serão apresentados nesta seção as diferenças entre as abordagens tendo como base a modelagem descrita na seção 2. Para começar, a distribuição de probabilidade conjunta da configuração $(\mathbf{y}, \mathbf{v}, \mathbf{h})$ é descrita por:

$$p(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = \frac{e^{-E(\mathbf{y},\mathbf{v},\mathbf{h};\boldsymbol{\theta})}}{\sum_{\mathbf{y},\mathbf{v},\mathbf{h}} e^{-E(\mathbf{y},\mathbf{v},\mathbf{h};\boldsymbol{\theta})}} \tag{12}$$

onde a função de energia é descrita por:

$$E(\mathbf{y}, \mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = -\sum_{i=1}^{m} \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} b_j h_j - \sum_{z=1}^{k} c_z y_z - \sum_{i,j=1}^{m,n} \frac{v_i}{\sigma^2} h_j w_{ij} - \sum_{z,j=1}^{k,n} y_z h_j u_{zj} \tag{13}$$

Das equações 12 e 13 as distribuições de probabilidade condicionais são obtidas. A distribuição de probabilidade da camada visível dado a camada oculta é idêntica a equação 7 da seção 2. Já a probabilidade da camada oculta dado a camada de rótulos e camada visível e a probabilidade da camada de rótulos dado a camada oculta, são descritas pelas equações 14 e 15, respectivamente.

$$p(h_j = 1|\mathbf{y}, \mathbf{v}; \boldsymbol{\theta}) = \phi(b_j + \sum_{z=1}^{k} y_z u_{zj} + \sum_{i=1}^{m} v_i w_{ij}) \tag{14}$$

$$p(y_z = 1|\mathbf{h}; \boldsymbol{\theta}) = \frac{e^{(c_z + \sum_{j=1}^{n} h_j u_{zj})}}{\sum_{l=1}^{k} e^{(c_l + \sum_{j=1}^{n} h_j u_{lj})}} \tag{15}$$

3.2. Treinando uma DRBM

O treinamento de uma DRBM segue os mesmos princípios do treinamento da RBM e por isso será apresentado de forma sucinta. O aprendizado dos pesos de conexão $\mathbf{W}$ e os vetores de *bias* $\mathbf{a}$ e $\mathbf{b}$, são obtidos pelas equações 9, 10 e 11. Todavia, para o processo de amostragem da camada escondida é utilizado a equação 14, como mostrado na Figura 5. Além disso, é necessário treinar os pesos de conexão $\mathbf{U}$ e o vetor de *bias* $\mathbf{c}$. O processo é similar e é descrito nas equações 16 e 17. A Figura 5 também mostra o processo de amostragem da camada de rótulos a partir da camada escondida utilizando a equação 15.

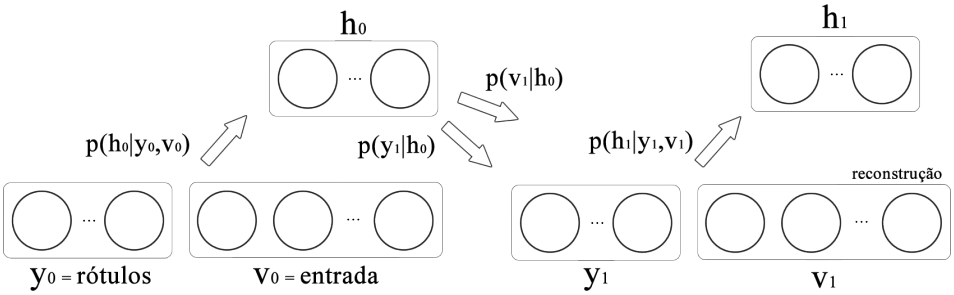


Figura 5: Processo de reconstrução do algoritmo CD para DRBM

$$\mathbf{U^{t+1}} = \mathbf{U^t} + \Delta\mathbf{U^t} \rightarrow \Delta\mathbf{W^t} = \eta(\mathbf{y_0}\mathbf{h_0^T} - \mathbf{y_1}\mathbf{h_1^T}) - \lambda\mathbf{U^t} + \alpha\Delta\mathbf{U^{t-1}} \tag{16}$$

$$\mathbf{c^{t+1}} = \mathbf{c^t} + \Delta\mathbf{c^t} \rightarrow \Delta\mathbf{c^t} = \eta(\mathbf{y_0} - \mathbf{y_1}) + \alpha\Delta\mathbf{c^{t-1}} \tag{17}$$

Por fim, dado uma amostra **x** relacionado a um conjunto de rótulos $e = (1, 2, ..., k)$, a classificação da DRBM pode ser obtida por meio da seguinte equação:

$$p(y^e|\mathbf{x}) = \frac{e^{-F(\mathbf{x},y^e)}}{\sum_{y^* \in \{1,2,...,k\}} e^{-F(\mathbf{x},\mathbf{y}^*)}} \tag{18}$$

O rótulo relacionado a amostra será aquele com maior probabilidade. A função $F(\mathbf{x}, y^e)$ é conhecida como "energia livre"e pode ser computada da seguinte maneira:

$$F(\mathbf{x}, \mathbf{y}) = c_e + \sum_{j=1}^{n} \varphi(\sum_{i=1}^{m} x_i w_{ij} + u_{yj} + b_j) \tag{19}$$

onde $\varphi(z) = log(1 + exp(z))$ e é conhecida como *softplus*. É importante salientar que o rótulo **y** é um vetor binário no qual apenas a posição e assume valor 1 quando **y** representa o rótulo e.

4. Resultados experimentais

Nesta seção a DRBM é aplicada para as cinco grandes bases de dados do repositório UCI [Lichman, 2013] apresentadas na Tabela 1. O principal objetivo deste experimento é verificar a eficácia do método para bases de dados que não sejam extraídas de imagens. Todos os procedimentos foram realizados no ambiente de desenvolvimento MATLAB utilizando a *toolbox* de rede neural nativa do ambiente, implementações próprias e a *toolbox DeebNet* [Keyvanrad e Homayounpour, 2014], que implementa algoritmos relacionados à máquina de Boltzmann restrita discriminativa. Os resultados obtidos pela DRBM são comparados com os seguintes algoritmos:

- Uma rede neural *feedforward* (FF) [Haykin, 2007] com duas camadas ocultas contendo 25 neurônios cada, com algoritmo de treinamento Levenberg-Marquardt *backpropagation* utilizando *mini-batch* de 100 partições, taxa de aprendizagem igual a 0.01 e 1000 iterações.
- Uma rede neural usando *Extreme Learning Machine* (ELM) [Huang et al., 2006] com 500 neurônios na camada oculta.
- Um K-*Nearest Neighbor* (KNN) [Keller et al., 1985] com k escolhido como $\sqrt{T}$, sendo T o tamanho da base de dados utilizada.

A DRBM utilizada possui 500 neurônios na camada oculta, utiliza *mini-batch* de 100 partições, 300 iterações e os parâmetros η, λ e α são os mesmos propostos por Hinton [2002] descritos nas seção 2.

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

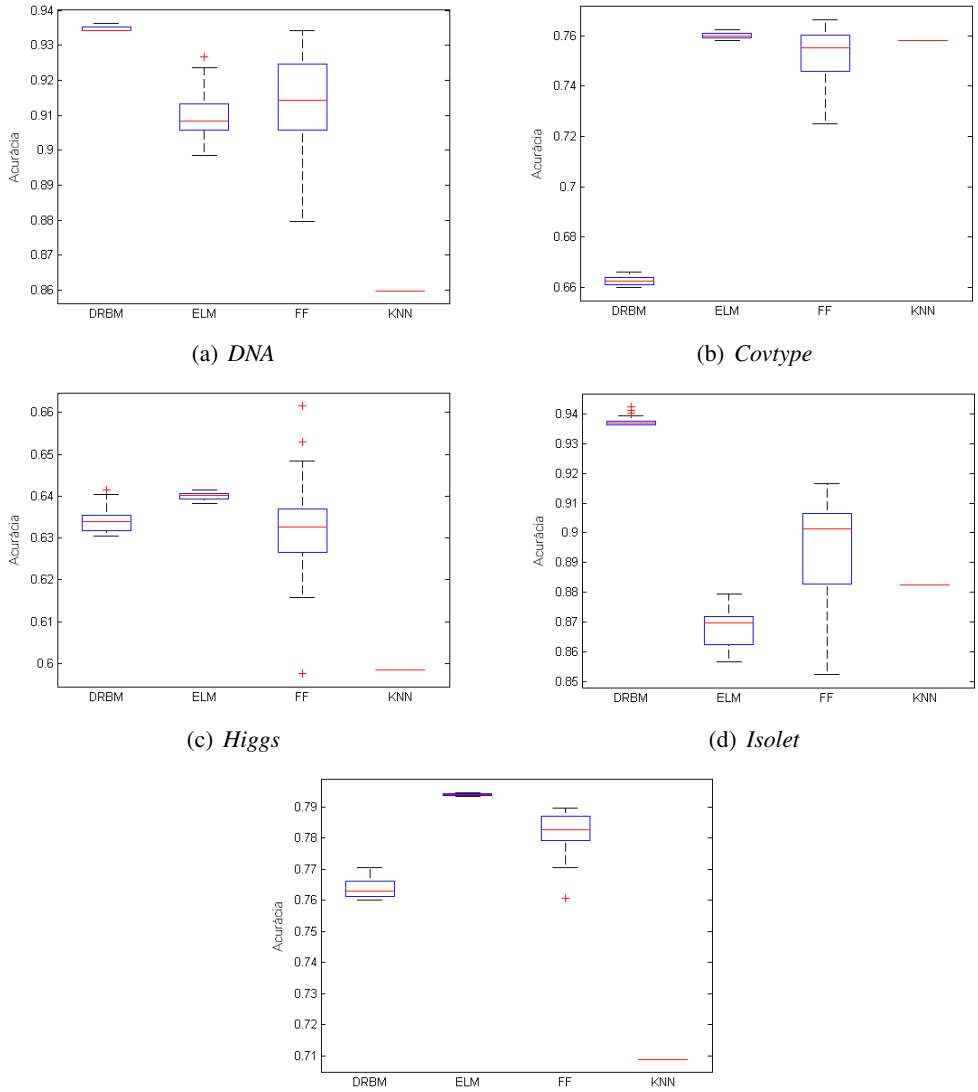
Tabela 1: Bases de dados utilizadas nos experimentos

Como mostrado na Tabela 1 as bases de dados utilizadas podem ser divididas em dois grupos: aquelas com muitas amostras, *covtype*, *higgs* e *susy*, sendo as duas últimas extratos de um milhão das bases originais, e aquelas com número elevado de atributos de entradas, *DNA* e *isolet*. Para executar os experimentos, cada base foi dividida em 70% para treinamento e 30% para testes.

Além disso, cada algoritmo, com exceção do KNN que não é estocástico, foi executado 30 vezes para cálculo de estatísticas do resultado. O desempenho de cada algoritmo, em termos de acurácia de classificação, é apresentado na Tabela 2 e nos *boxplots* da Figura 6.

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

Tabela 2: Desempenho dos classificadores para cada base de dados. Em negrito o classificador com melhor desempenho considerando a acurácia média



(a) *DNA*

(b) *Covtype*

(c) *Higgs*

(d) *Isolet*

(e) *Susy*

Figura 6: *Boxplots* do desempenho dos classificadores para cada base de dados

Por meio dos resultados apresentados na Tabela 2 e dos *boxplots* da Figura 6, é possível observar que a rede ELM e a DRBM se destacam entre os demais classificadores. Nas bases *covtype*, *higgs* e *susy*, a ELM apresentou os melhores resultados. Entretanto, quando o número de atributos da base aumentam significativamente, é o caso das bases DNA e *isolet*, o melhor classificador é a DRBM. A rede FF não foi o melhor classificador em nenhuma das execuções, mas mantém-se sempre próximo do melhor. Por fim, o KNN, por ser uma metodologia clássica e simples, obteve um resultado aceitável.

Focando nos resultados da DRBM, é possível afirmar que o algoritmo manteve-se competitivo nas bases de dados *susy* e *higgs*. Por outro lado, na base de dados *covtype* o método obteve o pior resultado comparado com as demais abordagens. Observando as bases de dados em que a DRBM foi o melhor classificador, *DNA* e *isolet*, é encontrado um padrão nas mesmas: valores elevados de atributos de entrada. Essa constatação vai ao encontro dos resultados obtidos no trabalho de Bu et al. [2015], na qual a RBM é utilizada para reduzir dimensionalidade de bases de dados com muitas entradas. Além disso, em classificação de imagens de reconhecimento de dígito, problema na qual a DRBM vem sendo utilizada, os dados de entrada possuem grande dimensionalidade, fato que também corrobora com os resultados obtidos neste trabalho.

5. Conclusão

O principal objetivo deste trabalho foi verificar a eficácia da máquina restrita de Boltzmann discriminativa quando utilizada para classificar grandes bases de dados que não fossem extraídas de imagens. Sendo assim, a DRBM foi aplicada para cinco grandes bases de dados do repositório UCI e o desempenho do algoritmo foi comparado com três classificadores amplamente utilizados na área: uma rede neural *feedforward*, uma rede *extreme learning machine* e um K-*nearest neighbors*. Os resultados obtidos por todos os classificadores mostraram que de maneira geral a DRBM foi competitiva nas bases utilizadas. Além disso, o desempenho do método apontou que o mesmo é uma boa escolha para bases de dados com alta dimensionalidade na entrada.

Referências

Bourlard, H. e Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4-5):291–294.

Bu, Y., Zhao, G., Luo, A.-l., Pan, J., e Chen, Y. (2015). Restricted Boltzmann machine: a non-linear substitute for PCA in spectral processing. *Astronomy & Astrophysics*, 576:A96.

Haykin, S. (2007). *Neural Networks: A Comprehensive Foundation (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.

Hinton, G. E. (2012). A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*, p. 599–619. Springer.

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

Hinton, G. E. e Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

Huang, G.-B., Zhu, Q.-Y., e Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501.

Ji, N., Zhang, J., Zhang, C., e Wang, L. (2014). Discriminative restricted Boltzmann machine for invariant pattern recognition with linear transformations. *Pattern Recognition Letters*, 45: 172–180.

Keller, J. M., Gray, M. R., e Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 15(4):580–585.

Keyvanrad, M. A. e Homayounpour, M. M. (2014). A brief survey on deep belief networks and introducing a new object oriented MATLAB toolbox (deebnet v2. 1). *arXiv preprint arXiv:1408.3264*.

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.

Lichman, M. (2013). UCI machine learning repository. URL `http://archive.ics.uci.edu/ml`.

Liu, K., Zhang, L. M., e Sun, Y. W. (2014). Deep Boltzmann machines aided design based on genetic algorithms. In *Applied Mechanics and Materials*, volume 568, p. 848–851. Trans Tech Publ.

Memisevic, R. e Hinton, G. E. (2010). Learning to represent spatial transformations with factored higher-order Boltzmann machines. *Neural Computation*, 22(6):1473–1492.

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.

Papa, J. P., Scheirer, W., e Cox, D. D. (2016). Fine-tuning deep belief networks using harmony search. *Applied Soft Computing*, 46:875–885.

Salama, M. A., Hassanien, A. E., e Fahmy, A. A. (2010). Deep belief network for clustering and classification of a continuous data. In *2010 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, p. 473–477.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. Technical report, DTIC Document.

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