

# Análise do impacto da redução de dimensionalidade em um algoritmo de aprendizado

Fabício Rodrigues Inácio  
Universidade Federal de Minas Gerais  
finacio@ufmg.br

Raquel Yuri da Silveira Aoki  
Universidade Federal de Minas Gerais  
raquel-aoki@ufmg.br

## ABSTRACT

Os algoritmos de aprendizado são técnicas que permitem que computadores sejam usados para aprender e/ou se especializar na resolução de determinado tipo de tarefa. Um dos usos mais comuns para esse tipo de algoritmo é realizar a classificação de dados. Para que essa classificação seja feita, os algoritmos avaliam os atributos dos dados e buscam uma correlação que permita agrupar dados semelhantes em uma mesma classe.

O número de atributos que devem ser avaliados para construir o modelo utilizado para classificar os dados está diretamente ligado à eficiência dos algoritmos de aprendizagem. Um grande número de atributos pode permitir que os algoritmos desenvolvam modelos mais robustos e específicos. Por outro lado, são mais susceptíveis a *overfitting* e o tempo de processamento pode crescer de forma rápida, tornando o uso do algoritmo inviável. Diante desse fato, alguns métodos tentam reduzir a dimensionalidade dos dados na tentativa de reduzir os recursos necessários para utilização dos algoritmos de aprendizagem.

Este artigo investiga o impacto da redução da dimensionalidade dos dados no desempenho de alguns algoritmos de aprendizado de máquina.

## Categories and Subject Descriptors

I.2 [ARTIFICIAL INTELLIGENCE]: Learning

## Keywords

Aprendizado de Máquina, redução de dimensionalidade, PCA, RBM, SVM.

## 1. INTRODUÇÃO

Uma das características que mais impactam no desempenho dos algoritmos de aprendizado é a dimensionalidade dos dados analisados. Diante desse fato, torna-se importante a tentativa de se reduzir o número de atributos avaliados pelos algoritmos. Para que essa redução seja feita, deve-se ter

um subconjunto de atributos que consiga representar os dados de forma equivalente ao conjunto completo. Atributos redundantes podem prejudicar o desempenho dos algoritmos de aprendizagem tanto na velocidade quanto na taxa de acerto, já que, devido à redundância, o número de atributos avaliados será maior e essas informações podem confundir o algoritmo, ao invés de auxiliá-lo na busca de um modelo correto para representar os dados [5].

Para avaliar o impacto que a redução de dimensionalidade dos dados pode causar no desempenho de algoritmos de aprendizagem, escolhemos um dos desafios disponíveis no site [www.kaggle.com](http://www.kaggle.com). O problema foi proposto por uma grande companhia de comércio eletrônico, o *Otto Group*, que vende milhões de itens e recebe milhares de novos produtos para serem comercializados diariamente. No entanto, devido a sua infraestrutura global e diversificada, muitas vezes produtos idênticos são classificados de maneiras distintas. Entretanto, a qualidade da análise de produtos depende da capacidade de agrupar com precisão os similares, e por isso, quanto melhor for a classificação, mais conhecimento pode ser gerado sobre sua gama de produtos. Portanto, o objetivo do desafio é obter um método que consiga classificar novos itens em uma das nove categorias em que o grupo divide seus produtos. Para isso, foram disponibilizadas duas bases de dados: o treino, que possui 61878 produtos descritos por 93 atributos e a classe a qual cada produto pertence; e a base de teste, que contém 144368 produtos com seus 93 atributos. A número de elementos em cada classe na base treino pode ser observada na Tabela 1.

Tabela 1: Número de elementos em cada classe

Classe	Nº de Elementos
1	1929
2	16122
3	8004
4	2691
5	2739
6	14135
7	2839
8	8464
9	4955

O modelo de classificação foi primeiramente desenvolvido com base nos dados do treino, e depois os elementos da base de teste foram atribuídos a uma classe a partir do modelo aprendido. A redução da dimensionalidade foi feita de três maneiras distintas: utilizando o PCA (*Principal Component*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$10.00.

*Analysis*) e o RBM (*Restricted Boltzmann Machines*). O método utilizado para separar as classes foi o SVM (*Support Vector Machines*).

No final, deseja-se comparar o desempenho das técnicas utilizados para classificar os itens do treino, no *cross-validation* e o que cada uma prediz para os itens da base de teste.

## 2. REFERENCIAL TEÓRICO

O classificador adotado nesse trabalho foi o SVM, um método de aprendizado de máquina supervisionado que aprende hiperplanos que separam as classes da base de dados. É considerado computacionalmente eficiente, pois funciona bem com amostras pequenas e grandes. O SVM foi utilizado para resolver um problema multi-classe, e neste caso, ele associa a cada uma das  $m$  classes um vetor de pesos e vícios. Geometricamente, é equivalente a associar um hiperplano para cada classe e atribuir o novo ponto  $x$  à classe onde o hiperplano está mais distante dele. Podem existir muitos hiperplanos que separam as classes, mas somente um deles é ótimo e maximiza a margem de separação das classes.

Uma constante  $C$  controla a magnitude da penalização quando ocorre um erro no treino. Se  $C$  é pequeno, uma grande fração de vetores suporte pode ser derivada, mas se o valor de  $C$  é alto, aumenta-se a chance de ocorrer *overfitting*, o que diminui a capacidade de generalização do SVM. Desta forma, ao acrescentar o termo de penalização dos erros, o problema do SVM é otimizar

$$\min \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^r \xi_i \right]$$

restrito a  $y(wx_i + b) > 1 - \xi_i$ , onde  $b$  é o vício,  $x$  é um ponto no hiperplano e  $w$  é perpendicular ao hiperplano. Se as classes se sobrepõem no espaço, então o termo  $C \sum_{i=1}^r \xi_i$  será grande e talvez o hiperplano não generalize tão bem. Para maiores detalhes sobre o SVM, ver [2].

A primeira técnica de redução de dimensionalidade utilizada foi o PCA, em português, Análise de Componentes Principais. Seu objetivo é explicar a estrutura de variância e covariância através da construção de combinações lineares dos atributos originais. As novas componentes obtidas são não correlacionadas entre si e a variabilidade da base completa é aproximada pela variabilidade da base reduzida, ou seja, a qualidade de aproximação depende do número de componentes deixadas na base reduzida. Mais informações em [6].

A RBM (*Restricted Boltzmann Machines*) é uma rede neural artificial estocástica que aprende uma distribuição de probabilidade a partir do treino. Seu objetivo é atribuir alta probabilidade para exemplos presentes no treino. Nesse trabalho, o RBM será utilizado apenas para reduzir a dimensionalidade dos dados e para tanto, a camada oculta deve ser do tamanho para o qual se deseja reduzir as bases de dados. Para que a camada oculta gere uma boa representação dos dados, deve-se executar o RBM normalmente e ao final, fazer uma multiplicação matricial do treino com a camada oculta. O SVM foi aplicado a essa base reduzida, constituída a partir do que se pode chamar de variáveis latentes, que não são variáveis observadas diretamente, mas inferidas a partir dos dados. Maiores detalhes em [3, 8] e [4].

Inicialmente, pensou-se em reduzir as bases através de uma terceira técnica, conhecida como t-SNE. Mas devido a dificuldade de sua utilização e a falta de informações sobre

sua implementação fez que com essa técnica fosse deixada de lado.

Uma maneira de calcular acurácia de um classificador multi-classes é através do Escore F1. O Escore F1 é calculado para cada classe considerando as medidas de precisão (fração de instâncias recuperadas que são relevantes) e revocação (fração de casos relevantes que são recuperados) do algoritmo de classificação [1]. Para um melhor entendimento do cálculo desses conceitos considere, por exemplo, que desejamos calcular a precisão e a revocação do classificador SVM para a classe 1. Considere como VP (verdadeiro positivo) os elementos que foram classificados na classe 1 e realmente pertencem a classe 1; FP (falso positivo) os elementos classificados na classe 1 mas que pertencem na verdade a outra classe e FN (falso negativo) os elementos classificados em outras classes mas que pertencem de fato a classe 1. Dessa forma, tem-se que a precisão e a revocação são calculada da seguinte maneira:

$$Precisao(classe1) = \frac{VP}{VP + FP}$$

$$Revocacao(classe1) = \frac{VP}{VP + FN}$$

O Escore F1 é então definido como:

$$F_1(classe1) = 2 * \frac{precisao(classe1) * revocacao(classe1)}{precisao(classe1) + revocacao(classe1)}$$

Para reduzir o risco de *overfitting* e tentar obter um modelo com maior capacidade de generalização, utilizamos a técnica conhecida como *k-fold cross-validation*. Ela propõe que o conjunto de dados seja dividido em  $k$  subconjuntos de tamanhos iguais denominados *folds*. Feito isso, o algoritmo de aprendizado deve ser executado  $k$  vezes. Em cada execução o algoritmo utiliza  $k-1$  folds para treinamento e 1 fold para a validação, sendo que em cada iteração, um fold diferente é usado na fase de validação.

Para executar os algoritmos de aprendizado e analisar os resultados dos experimentos feitos, utilizou-se o software estatístico R [7].

## 3. METODOLOGIA

Caso somente a base treino tivesse a dimensão reduzida, o SVM ajustado não seria capaz de fazer previsões na base de teste, pois eles teriam dimensões diferentes. Ao mesmo tempo, elas não poderiam ser reduzidas separadamente (mesmo que para a mesma dimensão) pois variáveis latentes distintas poderiam ser geradas. Portanto, na redução da dimensionalidade, tanto no PCA quanto no RBM, os atributos da base treino e teste foram concatenados e após aplicada a técnica de redução, elas foram separadas novamente. Além disso, como os atributos tinham escalas diferentes, para o funcionamento correto dos algoritmos de redução de dimensionalidade, a base de dados foi escalonada para a utilização do PCA e padronizadas no RBM.

O PCA foi feito utilizando a função *prcomp()*, nativa do ambiente R. A base pode ser reduzida de duas maneiras distintas com o PCA: na primeira, os componentes principais selecionados para constituir a base reduzida são aqueles que

possuem autovalores maiores que 1, resultado em 27 variáveis; no segundo, são selecionados os componentes principais necessários para que expliquem pelo menos 90% da variabilidade dos dados, resultando em 63 variáveis. Como deseja-se comparar os resultados obtidos utilizando diferentes métodos de redução de dimensionalidade com posterior utilização do SVM, essas duas formas de redução foram utilizadas.

O pacote *"deepnet"* foi utilizado para o ajuste do RBM através da função *rbm.train()*. Para ter resultados mais comparáveis, o RBM foi reduzido para as mesmas dimensionalidades dos dois métodos usados no PCA. Além disso, foram testadas duas épocas diferentes (10 e 100), que é o número de interações por amostra. Para facilitar a compreensão ao longo do texto, defini-se que RBM1 refere-se ao RBM ajustado com 10 épocas e o RBM2 ao de 100 épocas. Os demais parâmetros utilizados na função *rbm.train()* foram: *cd* = 1, *learningrate* = 0.1, *learningrate\_scale* = 0.5 e *momentum* = 1.

No total, são sete os resultados a serem comparados: um com o SVM aplicado a base completa e seis SVM's aplicados às bases reduzidas (2 PCA's e 4 RBM's).

Para o ajuste do SVM foi utilizado o pacote *"e1071"*, que em problemas multi-classe com  $m$  níveis, usa a abordagem "one-by-one", na qual  $m(m-1)/2$  classificadores binários são treinados. A função foi a *svm()* e os parâmetros foram os mesmos em todos os 7 ajustes: *gamma* = 0.9, *tolerance* = 0.9, *epsilon* = 0.5, *kernel* = "radial" e *cross* = 5.

Foi feita uma função no R para o cálculo do Escore F1. Para mais detalhes sobre a utilização dessas funções, indique o help do R[7].

## 4. RESULTADOS

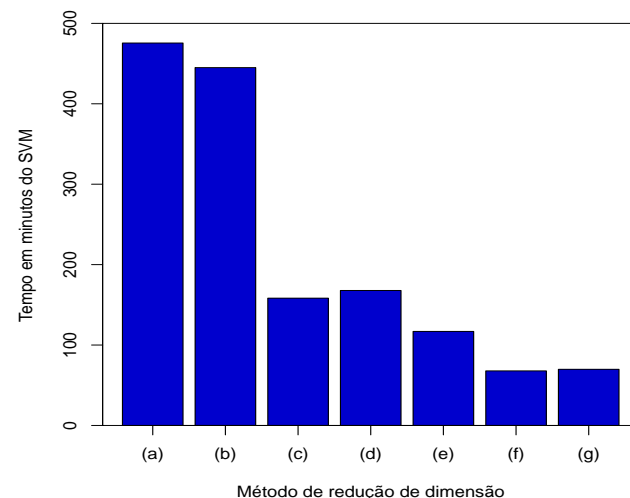
### 4.1 Acurácia

Na Tabela 2 são mostrados os resultados do Escore F1 para cada uma das técnicas. O modelo completo, que é o SVM com 93 colunas, foi o que teve maior acurácia em todas as classes. Considerando os SVM's nas bases reduzidas, o que apresentou melhor resultado foi o PCA com 63 colunas. Além do SVM sem redução, esse foi o único que apresentou acurácia maior que 90% em todas as classes, sendo que sua menor acurácia foi na classe 3 (93.68%) e a maior na 6 (99.85%). O PCA com 27 colunas também apresentou resultados satisfatórios, considerando a grande redução de dimensionalidade feita. Sua menor acurácia foi observada na classe 4 (75.68%), e foi bem superior a do RBM1 e RBM2 com 27 colunas (22.3% e 22.09% respectivamente), e do RBM1 e RBM2 com 63 colunas (73.37% e 72.23% respectivamente) nessa mesma classe. O RBM1 e RBM2 com 63 colunas apresentaram acurácia superior a 70% em todas as categorias, mas como já citado, tiveram desempenho geral pior que os PCA's. Os piores resultados foram observados nas classificações feitas nas bases reduzidas para 27 colunas através do RBM1 e RBM2.

Ainda na Tabela 2, observando as classes, nota-se que algumas apresentaram valores de acurácia maiores. Esse é o caso da classe 5, cuja menor acurácia foi 93.55% no SVM+RBM2 com 27 colunas. Já a classe 4 foi a que apresentou menor acurácia, com 22.09% no SVM+RBM2 com 27 colunas. Resumidamente, as classes 2, 5, 6, 7, 8 e 9 foram bem classificadas, apresentando altos valores de acurácia nas 7 repetições do SVM realizadas. Já as classes 1, 3 e 4 não foram tão bem classificadas, pois em alguns dos

SVM's chegam a apresentar acurácias menores que 45

### 4.2 Tempo de treinamento



**Figura 1: Tempo do SVM em minutos: (a) sem redução de dimensionalidade; (b) PCA 63 colunas; (c) RBM1 63 colunas; (d) RBM2 63 colunas; (e) PCA 27 colunas; (f) RBM1 27 colunas e (g) RBM2 27 colunas.**

A Figura 1 mostra os tempos de treinamento do SVM nas diferentes bases. A intenção é conseguir observar a partir desse gráfico, o impacto que a redução da dimensionalidade da base treino causa no tempo de treinamento. Os SVM's estão ordenados por quantidade de colunas, sendo que em (a) tem-se o SVM completo com 93 colunas e em (g) o SVM no RBM2 com 27 colunas. Observado isso, nota-se que o tempo decresce com a diminuição da dimensionalidade dos dados. O SVM completo levou 7.92 horas para rodar, enquanto o SVM no RBM1 com 27 colunas apenas 1.13 horas.

Considerando os SVM's com mesma dimensionalidade, nota-se que o SVM na base reduzida pelo PCA demorou mais que nas bases reduzidas pelo RBM. Nos SVM's feitos com 27 colunas, o PCA levou 1.95 horas, enquanto o RBM1 e RBM2 levaram 1.13 e 1.16 horas respectivamente. Já com 63 colunas, o tempo do SVM no PCA (7.41 horas) foi mais que o dobro dos RBM's (2.63 e 2.80 horas respectivamente). Entretanto, como já foi observado na análise da acurácia do SVM no treino, embora sejam mais rápidos para treinar, as bases com menor dimensionalidade possuem em geral acurácias menores. Além disso, os SVM's nas bases reduzidas utilizando RBM também tiveram acurácia pior que a dos SVM's nos PCA's com mesma dimensão.

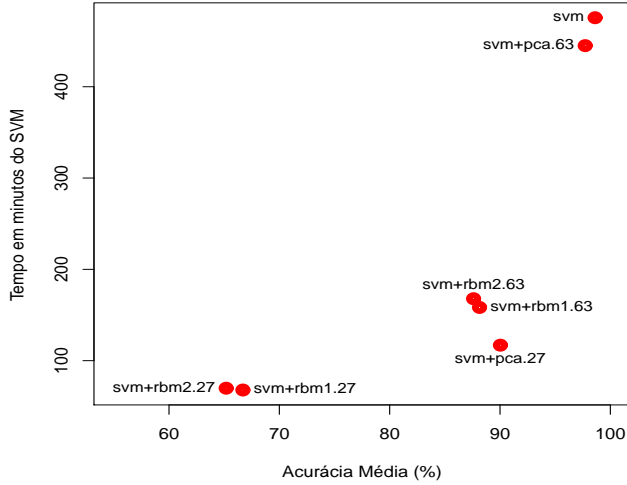
### 4.3 Relação entre tempo de treinamento e acurácia

Os gráficos de pontos mostrados a seguir foram feitos com o objetivo de entender melhor essa relação entre tempo e acurácia.

A Figura 2 mostra o tempo gasto no ajuste do SVM *versus* a acurácia média entre as classes. Os métodos podem ser divididos em três grupos: o primeiro é composto pelo SVM

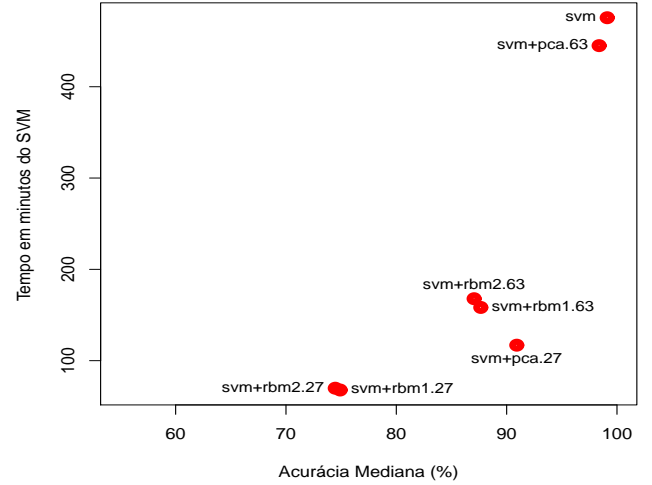
Tabela 2: Escore F1

Técnica	Dimensão	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6	Classe 7	Classe 8	Classe 9
SVM	93	99.12%	97.79%	96.27%	96.81%	99.62%	99.91%	98.9%	99.64%	99.5%
SVM+PCA	63	98.38%	96.36%	93.68%	94.53%	99.54%	99.85%	98.31%	99.53%	99.28%
SVM+RBM1	63	84.05%	87.62%	76.11%	73.37%	97.69%	97.81%	87.66%	95.45%	93.53%
SVM+RBM2	63	82.8%	87.05%	75.02%	72.23%	97.59%	97.71%	87.02%	95.53%	93.48%
SVM+PCA	27	89.44%	87.83%	76.33%	75.68%	98.21%	98.58%	90.92%	97.5%	95.82%
SVM+RBM1	27	45.75%	74.91%	39.98%	22.3%	93.05%	92.5%	65.3%	83.57%	82.91%
SVM+RBM2	27	38.91%	74.47%	41.6%	22.09%	92.64%	93.38%	61.01%	82.98%	79.56%

Figura 2: Tempo do SVM em minutos *versus* acurácia média em %.

aplicado às bases reduzidas pelo RBM com 27 colunas, que gastaram o menor tempo de ajuste, mas apresentaram os piores valores de acurácia média. O segundo grupo contém o SVM aplicado às bases do RBM reduzidas com 63 colunas e o SVM aplicado à base reduzida pelo PCA com 27 colunas, sendo que esse último demonstra ter um melhor custo benefício que os outros dois, pois tem uma acurácia média um pouco maior e necessita de menos tempo para ser ajustado. E por último, o grupo com maior acurácia média e que demanda mais tempo de ajuste, composto pelo SVM completo e pelo SVM aplicado a base reduzida pelo PCA com 63 colunas.

Como a média é fortemente influenciada por valores extremos e na Tabela 2 foram observados alguns valores muito baixos de acurácia em algumas classes, foi feita a Figura 3 com a acurácia mediana em porcentagem. A mediana foi escolhida por ser uma medida de tendência central menos influenciada por valores extremos. Os grupos 2 e 3 citados anteriormente não apresentaram grandes mudanças, mas o grupo 1, composto pelo SVM aplicado as bases reduzidas pelo RBM com 27 colunas, apresentou diferenças significativas. Antes seus valores de acurácia média oscilavam entre 65 e 67%, agora as acurácias medianas oscilam entre 74 e 75%, mostrando que antes a acurácia média estava subestimando esses classificadores. Ainda assim, seus resultados são inferiores com relação aos demais grupos.

Figura 3: Tempo do SVM em minutos *versus* acurácia mediana em %.

Comparando os resultados do RBM1 ajustado com 10 épocas e do RBM2 com 100 épocas, com auxílio das Figuras 1, 2 e 3, observa-se que as bases reduzidas para a mesma dimensão possuem tempos e acurácias parecidas, sendo que os resultados do RBM2 foram ligeiramente piores considerando a acurácia média e mediana, e o ajuste do SVM foi alguns minutos mais lento que o RBM1.

#### 4.4 Cross-Validation

O *cross-validation* foi rodado com 5 folds e a Tabela 3 mostra seu resultado médio. Até o momento, as análises indicavam que os melhores SVM's eram o completo e o aplicado ao PCA com 63 colunas. Entretanto, a Tabela 3 mostra que, muito provavelmente, esses modelos tiveram *overfitting*. Os modelos SVM completo e do SVM no PCA com 63 colunas tem taxa de acurácia de apenas 43.64% e 45.83% respectivamente, bem inferior aos demais modelos, todos superiores a 70%. Observe ainda que os SVM's rodados nas bases reduzidas pelo RBM são os que apresentam maiores resultados.

#### 4.5 Predições

Até o momento, todas as análises foram voltadas para a adequação dos SVM's ajustados à base treino. A partir de agora, serão analisadas as predições feita a partir de cada um desses SVM's.

A Tabela 4 mostra a proporção de vezes que duas técnicas

**Tabela 4:** Tabela com a proporção de concordância e discordância entre as técnicas utilizadas na base de teste.

Concor.	Discor.	Dimensão 93	Dimensão 63			Dimensão 27		
		SVM	SVM+PCA	SVM+RBM1	SVM+RBM2	SVM+PCA	SVM+RBM1	SVM+RBM2
SVM	93	1.00	0.05	0.45	0.46	0.41	0.53	0.53
SVM+PCA	63	0.95	1.00	0.43	0.44	0.38	0.51	0.51
SVM+RBM1		0.55	0.57	1.00	0.08	0.16	0.21	0.21
SVM+RBM2		0.54	0.56	0.92	1.00	0.17	0.20	0.21
SVM+PCA	27	0.59	0.62	0.84	0.83	1.00	0.27	0.27
SVM+RBM1		0.47	0.49	0.79	0.80	0.73	1.00	0.12
SVM+RBM2		0.47	0.49	0.79	0.79	0.73	0.88	1.00

**Tabela 3: Resultado do Cross-Validation**

Técnica	Dimensão	Resultado <i>Cross-Validation</i>
SVM	93	43.64
SVM+PCA	63	45.83
SVM+RBM1	63	71.79
SVM+RBM2	63	71.97
SVM+PCA	27	70.09
SVM+RBM1	27	72.35
SVM+RBM2	27	71.93

concordaram ou discordaram na classificação da base teste. Relembra-se que a base teste foi reduzida juntamente com o treino. A diagonal principal é a concordância do classificador com ele mesmo, por isso é 1. Abaixo da diagonal principal está a proporção de concordância entre as técnicas, e acima da diagonal está a proporção de discordância. Observe que a soma da proporção de concordantes e discordantes entre duas técnicas deve ser sempre 1. Por exemplo, o SVM aplicado ao RBM1 e o SVM aplicado ao RBM2, ambos com 27 colunas, concordaram na classificação da base teste com uma proporção igual a 0.88 e discordam com uma proporção de 0.12, somando esses dois valores obtém-se 1. Além disso, a proporção também pode ser interpretada como um valor percentual. Note que se duas técnicas tem uma proporção de concordância igual a 0.73, é o mesmo que dizer que elas concordam 73% das vezes. A tabela mostra que as predições do SVM completo e o SVM aplicado ao PCA com 63 colunas concordam em 95% dos casos. Entretanto, ambos concordam em menos de 62% dos casos com as demais técnicas de redução. Essas outras técnicas tendem a concordar mais entre si, discordando em no máximo 27% das vezes. O SVM aplicado ao RBM1 e RBM2 com 63 colunas, por exemplo, concordam em 92% das vezes. O SVM aplicado ao PCA com 27 colunas concorda mais com os SVM's aplicados as bases reduzidas pelo RBM1 e 2 com 63 e 27 colunas do que com o PCA de 63 colunas.

No geral, as técnicas que mais concordaram entre si foram o SVM completo e o SVM aplicado ao PCA com 63 colunas (95%), e as que menos concordaram foram o SVM completo e os SVM's aplicados às bases reduzidas pelo RBM1 e RBM2 com 27 colunas (47%).

A Figura 4 é composta de vários gráficos de barras, e cada um mostra a proporção das classes nas predições feitas por cada uma das técnicas aplicadas a base de teste. Note que os eixos verticais dos gráficos variam.

O gráfico da classe 1 mostra que as técnicas (a) (SVM completo) e (b) (SVM aplicado ao PCA com 63 colunas) classifi-

cou bem menos elementos nessa categoria que as demais técnicas, sendo que as técnicas que mais classificaram elementos nessa categoria foram a (c) (SVM aplicado ao RBM1 com 63 colunas) e (d) (SVM aplicado ao RBM2 com 63 colunas). Novamente, na classe 2 observa-se que (a) e (b) classificaram menos elementos nessa categoria que as demais, e dessa vez as técnicas que mais classificaram elementos na classe 2 foram (f) (SVM aplicado ao RBM1 com 27 colunas) e (g) (SVM aplicado ao RBM1 com 27 colunas). A classe 3 foi mais presente nas predições de (e) (SVM aplicado ao PCA com 27 colunas) e menos nas de (a). A classe 4 esteve presente com praticamente as mesmas proporções em (a), (b), (f) e (g), e foi observada com maior proporção em (e). As técnicas (a) e (b) foram as únicas que diferiram das demais quanto a proporção de elementos preditos na classe 5, predizendo um número menor que as outras. Na classe 6 o fenômeno foi o inverso: as técnicas (a) e (b) classificaram quase o dobro de elementos da base teste nessa categoria que as técnicas (c), (d), (e) e (f). O método que mais classificou elementos na categoria 7 foi (f) e o que menos classificou foi (a). As classes 8 e 9 foram classificadas de maneira semelhante a classe 7.

De forma geral, as técnicas (a) e (b) subestimaram a proporção de todas as classes com relação as demais, com exceção da classe 6, onde essas duas técnicas classificaram cerca de 70% dos elementos da base teste, enquanto as outras técnicas classificaram em torno de 30%. Esse comportamento pode ser explicado em parte pelo *overfitting*.

Nota-se que a proporção de elementos classificados em cada uma das categorias pelas técnicas (c) e (d) é muito semelhante, assim como entre (f) e (g). Como essas técnicas são RBM's, pode-se inferir que o número de épocas não muda significativamente a maneira com que a base é reduzida, e consequentemente, o SVM encontra partições parecidas. Entre os RBM's, a mudança do tamanho da base acarreta um impacto bem maior que o número de épocas.

A técnica (e) tem um comportamento bem distinto de (b), que corresponde a outra base reduzida pelo PCA. As proporções de elementos classificados pelo SVM aplicado ao PCA com 27 colunas tende a seguir uma tendência parecida com a dos SVM's aplicados ao RBM. Isso indica que ele não sofreu *overfitting* como o SVM aplicado ao PCA com 63 colunas.

## 5. CONCLUSÃO

O objetivo desse trabalho era fazer um comparativo que o impacto da redução de dimensionalidade ocasionava no SVM, e isso foi feito analisando três tamanhos diferentes de dimensionalidade (93, 63 e 27 colunas) e duas técnicas

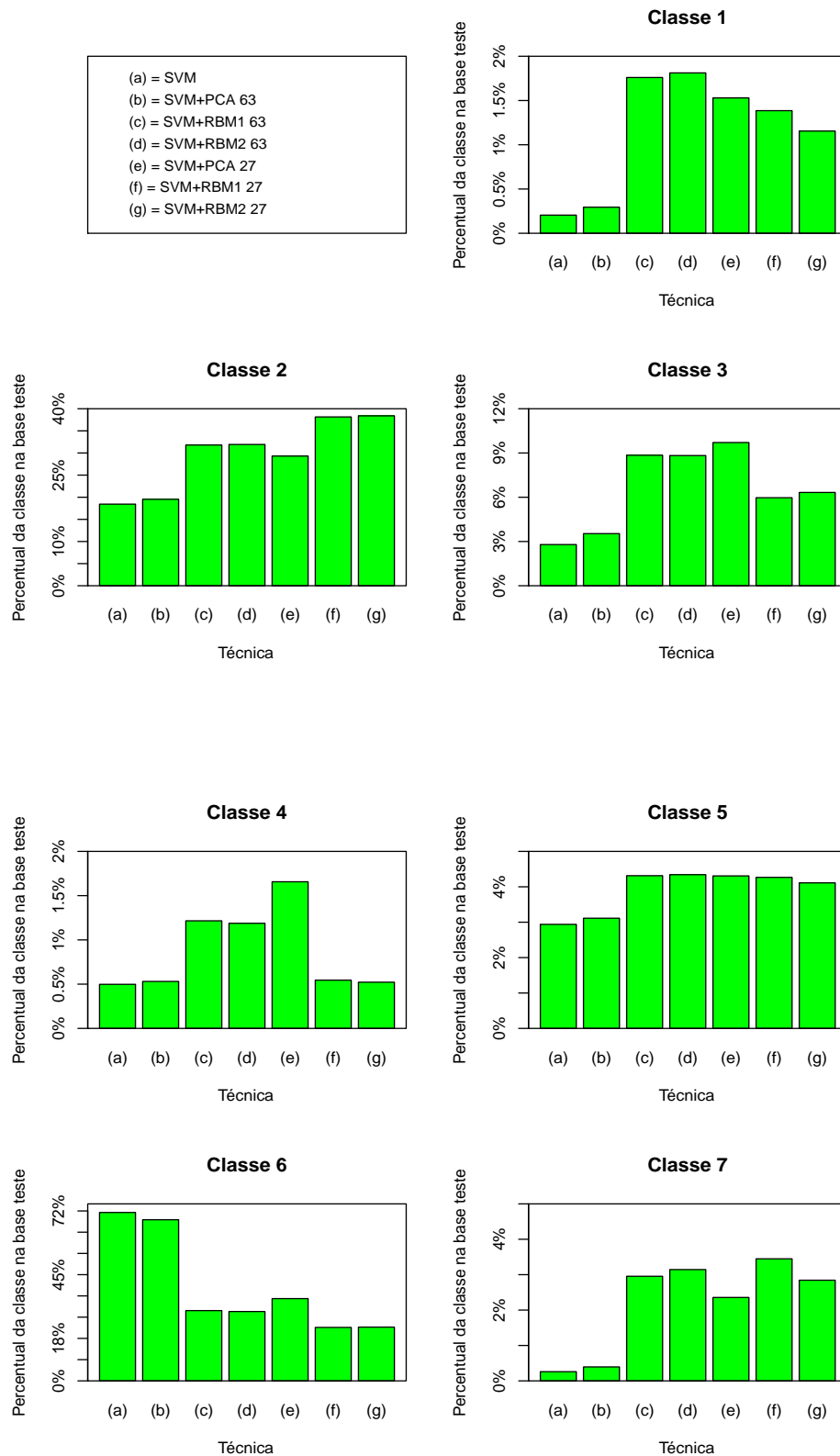


Figura 4: Percentual das classes na base teste em cada uma das técnicas estudadas.

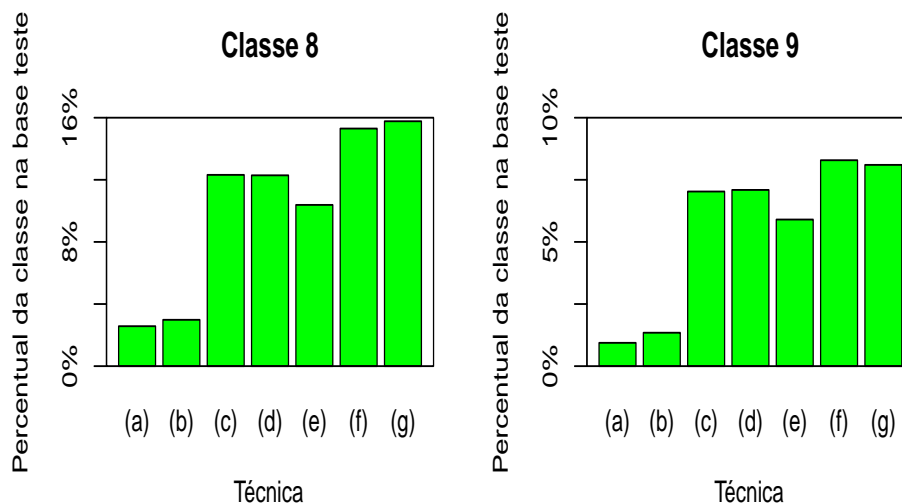


Figura 5: Continuação do percentual das classes na base teste em cada uma das técnicas estudadas.

diferentes de redução de dimensionalidade (PCA e RBM).

Analisando somente os resultados obtidos na base treino, a acurácia obtida pelo Escore F1 mostrou que o SVM diminuía a medida que a dimensão dos dados diminuía. Em alguns casos, como o PCA, essa redução não foi tão impactante para o resultado. Os Escores F1 do PCA com 27 colunas foram inferiores aos de 63 colunas, mas ainda assim, foram valores relativamente grandes, próximos aos do RBM com 63 colunas. Já no RBM, a redução de dimensionalidade de 63 colunas para 27 causou mais impacto e passou a ter uma acurácia inferior a 50% em algumas classes.

O tempo de treinamento dos SVM's é diretamente proporcional a dimensionalidade dos dados, entretanto, foi observado que as bases reduzidas pelos RBM's foram mais rápidas que as reduzidas pelos PCA's dentro de uma mesma dimensão.

Foram feitos gráficos do tempo de treinamento *versus* a média e mediana dos Escores F1 para facilitar a compreensão da relação entre os dois. Observou-se que a relação entre elas é quase exponencial: quanto mais tempo demorar o treinamento, maior será a acurácia média ou mediana.

Entretanto, ao analisar o resultado do *cross-validation* do SVM nota-se que os ajustes que melhor pareciam se adequar aos dados (SVM completo e o SVM aplicado a base reduzida com PCA para 63 colunas) foram os com menor taxa de acurácia. Acredita-se que esses dois modelos sofreram *overfitting*. Todas as demais taxas de acurácia foram superiores a 70%.

Mesmo após a conclusão de que houve *overfitting*, a predição foi feita para todas as técnicas, pois o objetivo do trabalho não é chegar ao melhor resultado, mas compara-los. Foi observada que os dois SVM's com *overfitting* predizem de maneira muito semelhante, concordando em 95% das vezes e concordando em aproximadamente 55% com as demais técnicas. Já as demais técnicas que apresentaram uma taxa de acurácia maior no *cross-validation*, concordaram mais entre si, indicando que predizem de maneira semelhante na maioria dos casos.

Após uma análise gráfica comparando a proporção das classes nos valores preditos pelas técnicas na base teste, novamente foi possível observar que as predições do SVM completo e o SVM aplicado a base reduzida com PCA para 63 colunas realmente diferem das demais e são muito semelhantes entre si. A principal característica que pode ser observada, é que essas duas técnicas tendem a classificar muitos elementos na classe 6 e subestimam todas as outras classes. Tal fato pode ser em parte explicado pelo fato que a classe 6 é a segunda mais frequente na base treino, e devido ao *overfitting*, essas técnicas predisseram mais elementos nessa categoria. A predição feita com o SVM aplicado as bases reduzidas com RBM para uma mesma dimensão é semelhante na maior parte das categorias. O SVM aplicado ao PCA com 27 colunas apresenta um comportamento distinto das demais técnicas, embora classifique de maneira mais parecida com os SVM's aplicados aos RBM's do que a do SVM completo e do SVM aplicado ao PCA com 63 colunas.

## 6. REFERENCES

- [1] R. Baeza-Yates and B. Ribeiro-Neto. *Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca*. Bookman, 2 edition, 2013.
- [2] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [3] G. M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(6):1335–1343, 2004.
- [4] G. Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.
- [5] K. Kira and L. A. Rendell. The feature selection problem: Traditional methods and a new algorithm. In *AAAI*, volume 2, pages 129–134, 1992.
- [6] S. A. Mingoti. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG, 2005.

- [7] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.
- [8] R. Salakhutdinov, A. Mnih, and G. Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM, 2007.