

## ESTUDO SOBRE TÉCNICAS DE REGULARIZAÇÃO DA MATRIZ COVARIÂNCIA NO PROCESSO DE CLASSIFICAÇÃO DE DADOS EM ALTA DIMENSIONALIDADE

MAURO ERBERT<sup>1,2</sup>  
VITOR HAERTEL<sup>2</sup>

<sup>1</sup> ULBRA - Universidade Luterana do Brasil  
Rua Martinho Lutero nº 301, Bairro Gonçalves, Cachoeira do Sul – RS, 96.501-000, Brasil  
merbert@uol.com.br

<sup>2</sup> UFRGS - Universidade Federal do Rio Grande do Sul  
CEPSRM - Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia  
Cx. Postal 15044- Porto Alegre – RS, 91.501-970, Brasil  
victor.haertel@vortex.ufrgs.br

**Abstract** - High dimensional image data that are now becoming available, offer new possibilities in image classification, specially when dealing with classes that present very similar spectral response. High dimensional image data poses, however, the problem of obtaining accurate estimates of the parameters required by statistical classifiers. This problem is caused by the small number of training samples usually available in real world conditions. Different approaches have been proposed in the literature aiming to mitigate this problem. One approach involves the techniques of regularization of the covariance matrix. This study investigates the applications of one regularization technique to high dimensional image data. Tests are performed using AVIRIS data, covering agricultural fields, and the results are presented and discussed

**Keywords:** sensoriamento remoto, alta dimensionalidade, análise discriminante regularizada, método supervisionado, reconhecimento de padrões

### 1. INTRODUÇÃO

A recente disponibilidade de dados em alta dimensionalidade, para aplicações em sensoriamento remoto, abre novas possibilidades, até então não disponíveis com a utilização de dados tradicionais em baixa dimensionalidade como dados LandSat-TM e SPOT, por exemplo. Dados com resolução espectral muito alta permitem a separação de classes espectralmente muito semelhantes. Pode-se mostrar (Fukunaga (1990)), que com dados em alta dimensionalidade, é possível separar classes que possuem idêntica resposta espectral, isto é, vetores média iguais, desde que as matrizes covariância difiram suficientemente entre si. O emprego de dados em alta dimensionalidade apresenta entretanto, em situações reais, o problema da estimação dos parâmetros requeridos por classificadores estatísticos, em especial, a estimação da matriz covariância. Sempre que o número de amostras de treinamento é pequeno, se comparado com a dimensionalidade dos dados, as estimativas dos parâmetros torna-se incerta, resultando numa degradação na performance do classificador. Por outro lado, é conhecida a limitação quanto ao número de amostras de treinamento em situações reais, fato este, que em princípio coloca severas limitações na utilização de dados em alta dimensionalidade, conhecidos como dados *hiperespectrais*. Na tentativa de minorar os efeitos causados pelo número insuficiente de amostras de treinamento, vários pesquisadores vem

investigando abordagens que permitam uma alternativa. De um modo geral, diferentes abordagens vem sendo propostas regularmente na literatura, sendo as principais (1) redução da dimensionalidade dos dados, para cada caso em particular, (2) aumento do número de amostras de treinamento por meio da introdução de novas amostras, conhecidas como “amostras não rotuladas”, obtidas com o emprego de algoritmos do tipo EM (expectation maximization) e (3) melhoria na estimação dos parâmetros, especialmente da matriz covariância, pelo emprego de técnicas conhecidas como “técnicas de regularização”. Este estudo investiga a utilização das técnicas de regularização da matriz covariância, originalmente propostas por Friedman (1989) com vistas à classificação de dados simulados (dados sintéticos) e dados relativos a enologia, ao processo de classificação de imagens hiperespectrais. Os experimentos foram desenvolvidos utilizando-se dados do sensor hiperespectral AVIRIS, e a cena utilizada cobre uma área agrícola localizada no Estado de Indiana – EUA. Esta área foi idealizada para fins de testes de dados em alta dimensionalidade, sendo coberta por culturas agrícolas que apresentam um comportamento espectral muito semelhante entre si, fato este que as torna quase que não separáveis a partir de dados tradicionais em baixa dimensionalidade. Esta área possui também informações relativas à “verdade terrestre” extremamente detalhadas, os quais permitem uma análise precisa dos resultados dos testes desenvolvidos.

## 2. MÉTODOS

Um dos classificadores mais utilizados no processo de classificação de imagens digitais em sensoriamento remoto é o da Máxima Verossimilhança (MV), e a respectiva função decisão apresenta a seguinte forma geral:

$$G_i(X) = p(X | \mathbf{w}_i)P(\mathbf{w}_i) \quad (1)$$

$$X \in \mathbf{w}_i \text{ se } G_i(X) > G_j(X), \text{ para todo } j \neq i \quad (2)$$

onde,  $X$  é o vetor que representa um *pixel* individual a ser classificado,  $p(X | \mathbf{w}_i)$  representa a função densidade de probabilidade associada à classe  $\mathbf{w}_i$  e  $P(\mathbf{w}_i)$  representa a probabilidade *a priori* associada à classe.

A utilização prática do classificador MV, requer a especificação da função densidade de probabilidade ( $p(X | \mathbf{w}_i)$ ). A experiência prática indica a função densidade de probabilidade Gaussiana como a mais adequada aos dados de sensoriamento remoto. A justificativa para este fato, é fornecida pelo Teorema Central do Limite. Neste caso, o classificador passa a denominar-se de Máxima Verossimilhança Gaussiana (MVG) e a função decisão (1), após algumas simplificações bem conhecidas adquire a forma:

$$G_i(X) = \frac{1}{\sigma_i} \exp \left\{ -\frac{1}{2\sigma_i} \left[ (X - \mathbf{m})^T \Sigma_i^{-1} (X - \mathbf{m}) + 2 \ln(P(\mathbf{w}_i)) \right] \right\}, \quad i=1, \dots, \text{número de classes (k)} \quad (3)$$

onde  $\mathbf{m}$  é o vetor média associada à classe  $\mathbf{w}_i$ , e  $\Sigma_i$  é a matriz covariância associada à classe  $\mathbf{w}_i$ . Os parâmetros  $\mathbf{m}$  e  $\Sigma_i$  são normalmente estimados a partir das amostras de treinamento disponíveis, por meio dos estimadores:

$$\hat{\mathbf{m}} = \frac{1}{N_i} \sum_{i=1}^{N_i} X_i \quad \hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{i=1}^{N_i} (X_i - \hat{\mathbf{m}})(X_i - \hat{\mathbf{m}})^T \quad (4)$$

onde,  $N_i$  é o número de amostras de treinamento disponíveis na classe  $\mathbf{w}_i$ .

Note-se que a função decisão (3) é quadrática. O processo de classificação que a implementa recebe a designação genérica de *Análise Discriminante Quadrática (ADQ)*. Em alguns casos, é aceitável assumir que todas as classes compartilham de uma mesma matriz covariância comum ( $\Sigma_1 = \Sigma_2 = \dots = \Sigma$ ). Este processo de classificação é genericamente denominado *Análise Discriminante Linear (ADL)*, e a matriz covariância comum pode ser estimada por:

$$\hat{\Sigma} = \frac{1}{N} \sum_{i=1}^k N_i \hat{\Sigma}_i \quad N = \sum_{i=1}^k N_i \quad (5)$$

Deste modo, o termo quadrático torna-se comum para todas as classes, então a função decisão (3) simplifica-se, tornando-se uma função linear em  $X$ :

$$G_i(X) = -\frac{1}{2}(X - \hat{\mathbf{m}})^T \hat{\Sigma}^{-1}(X - \mathbf{m}) + \ln(P(\mathbf{w}_i)) \quad (6)$$

O número de parâmetros a serem estimados, especialmente no caso da matriz covariância, cresce rapidamente com a dimensionalidade dos dados. Assim, por exemplo, no caso de dados LandSat-TM (7 bandas) o número de parâmetros a serem estimados é igual a 28. No caso do sensor AVIRIS (224 bandas) este número cresce para 25.200. A degradação na acurácia dos resultados da classificação com o aumento da dimensionalidade dos dados, mantendo constante o número de amostras de treinamento, é conhecida como *fenômeno de Hughes* ou a *maldição da dimensionalidade* (Tadjudin (1998)). Este fenômeno ocorre quando em um primeiro momento o acréscimo de novas bandas espectrais, faz com que a acurácia da classificação tenda a aumentar alcançando um máximo para a acurácia. A partir daí, um acréscimo no número de bandas espectrais passa a causar um decréscimo na acurácia fornecida pelo processo de classificação. Em outras palavras, o acréscimo de novos dados, ao invés de incrementar a acurácia do processo de classificação, tende a produzir, a partir de um momento, o efeito inverso. Este fato é devido à degradação da estimação da matriz covariância. Este problema, bem como possíveis alternativas para mitigá-lo, vem sendo objeto de estudo por parte de vários pesquisadores, como Shahshahani (1994), Hsieh (1998), Tadjudin (1998), Friedman (1989), Aebhard (1993) e Cortijo F.J. (1999). A técnica proposta por Friedman (1989), faz uso de dois parâmetros ( $\lambda$  e  $\gamma$ ) que variam no intervalo [0,1]. Através da utilização de diferentes valores numéricos para  $\lambda$  e  $\gamma$ , obtém-se situações intermediárias entre os métodos *ADQ* e *ADL*. Estas combinações têm por objetivo combinar em diferentes graus tanto a estimativa da matriz covariância comum das classes, quanto da matriz covariância de cada classe individual, resultando em uma estimativa mais confiável denominada *matriz covariância regularizada* ( $\hat{\Sigma}_i(\mathbf{I}, \mathbf{g})$ ).

Como o número de amostras de treinamento é limitado, as estimativas não-tendenciosas  $\hat{\Sigma}_i$  (4) para cada classe, podem não ser confiáveis. Nestes casos, a combinação linear da matriz covariância  $\hat{\Sigma}_i$  de cada classe e a matriz covariância  $\hat{\Sigma}$  comum entre as classes poderá produzir uma melhor estimativa da estrutura básica das classes, e

consequentemente um melhor resultado na classificação. Este processo se constitui no primeiro passo da regularização, e é expresso por:

$$\hat{\Sigma}_i(\mathbf{I}) = \frac{(1 - \mathbf{I})\hat{Q}_i + \mathbf{I}\hat{Q}}{N_i(\mathbf{I})}, \quad 0 \leq \mathbf{I} \leq 1 \quad (8)$$

para:

$$N_i(\mathbf{I}) = (1 - \mathbf{I})N_i + \mathbf{I}N \quad (9)$$

$$\hat{Q} = \sum_{i=1}^k \hat{Q}_i \quad \hat{Q}_i = N_i \hat{\Sigma}_i \quad (10)$$

onde,  $\hat{\Sigma}_i$  representa a matriz covariância própria de cada classe individual, estimada em (5),  $\hat{\Sigma}_i(\mathbf{I})$  representa a matriz covariância regularizada pelo parâmetro  $\mathbf{I}$ ,  $N$  é o somatório das amostras ( $N_i$ ) de todas as classes,  $\hat{Q}_i$  a matriz covariância de cada classe  $\hat{\Sigma}_i$  ponderada pelo tamanho da amostra de treinamento correspondente, e  $\hat{Q}$  é o somatório das estimativas  $\hat{Q}_i$ . A equação (10) tem o objetivo de ajustar o grau de influência do tamanho das amostras em cada classe, levando em consideração o número total de amostras e/ou o tamanho individual de amostras de cada classe. A equação (8) implementa um processo semelhante, porém com o termo  $\hat{Q}_i$ , resultando no primeiro passo  $\hat{\Sigma}_i(\mathbf{I})$  de regularização da matriz covariância. O parâmetro  $\mathbf{I}$  controla portanto a contribuição da estimativa da matriz covariância  $\hat{\Sigma}_i$  e da estimativa da matriz covariância comum  $\hat{\Sigma}$ .

Este primeiro procedimento de regularização da matriz covariância resulta portanto em um classificador intermediário entre os métodos *ADL* e *ADQ*, e normalmente já produz uma melhoria significativa no resultado do processo de classificação, especialmente quando o tamanho de amostras das classes individuais é pequeno. A regularização introduzida pela equação (8) ainda é entretanto limitada, por duas razões: (1) pode não realizar uma regularização suficiente, se o número de amostras é inferior ou comparável à dimensionalidade e (2) ajustar a matriz covariância de cada classe por meio da matriz covariância comum pode não ser a forma mais efetiva para melhorar as estimativas  $\hat{\Sigma}_i$ , pela razão de que em alguns casos este processo pode introduzir tendências severas (Friedman (1989)).

Para corrigir este problema, Friedman (1989) sugere um segundo passo no processo de regularização, o qual tem por objetivo corrigir o fato bem conhecido de que quando o número de amostras de treinamento é pequeno, os valores estimados para os autovalores da matriz covariância tornam-se tendenciosos: os autovalores menores são estimados ainda menores e os maiores ainda maiores. Este segundo passo de regularização é implementado pela equação (11). O efeito deste processo será diminuir os autovalores maiores e aumentar os autovalores menores.

$$\hat{\Sigma}_i(\mathbf{I}, \mathbf{g}) = (1 - \mathbf{g})\hat{\Sigma}_i(\mathbf{I}) + \frac{\mathbf{g}}{n} \text{tr}[\hat{\Sigma}_i(\mathbf{I})] \mathbf{I}, \quad \begin{array}{l} 0 \leq \mathbf{I} \leq 1 \\ 0 \leq \mathbf{g} \leq 1 \end{array} \quad (11)$$

onde:

$\text{tr}[\hat{\Sigma}_i(\mathbf{I})]$ : é o traço de  $\hat{\Sigma}_i(\mathbf{I})$ ;

$\mathbf{I}$  : é a matriz identidade.

Nota-se que  $tr[\hat{\Sigma}_i(\mathbf{I})]$  é igual à soma dos autovalores de  $\hat{\Sigma}_i(\mathbf{I})$ , e portanto o quociente  $tr[\hat{\Sigma}_i(\mathbf{I})]/n$  representa a média dos autovalores de  $\hat{\Sigma}_i(\mathbf{I})$ .

Empregando  $\hat{\Sigma}_i(\mathbf{I}, \mathbf{g})$  no lugar da  $\hat{\Sigma}_i$  na função decisão (4), a função decisão regularizada torna-se:

$$G_i(X) = -(X - \mathbf{m}_i)^T \hat{\Sigma}_i^{-1}(\mathbf{I}, \mathbf{g})(X - \mathbf{m}_i) - \ln |\hat{\Sigma}_i(\mathbf{I}, \mathbf{g})| + P(\mathbf{w}_i) \quad (12)$$

Aeberhard (1993), salienta não existir um processo que permita definir quais as melhores configurações de parâmetros de  $\lambda$  e  $\gamma$ , para que se tenha uma probabilidade de classificação correta (*PCC*) maximizada. A avaliação dos resultados do classificador é realizada experimentalmente, com um conjunto de pares de valores numéricos ( $\lambda, \gamma$ ), e então são selecionados os valores onde o resultado tenha fornecido a melhor acurácia.

### 3. EXPERIMENTOS

Os dados espectrais utilizados neste estudo são provenientes do sensor hiperespectral AVIRIS. O sensor AVIRIS dispõe de 224 bandas espectrais, utiliza uma plataforma aerotransportada, e é considerado um sensor experimental. A cena utilizada cobre uma área experimental, localizada no Estado de Indiana, Estados Unidos. Esta área está ocupada por culturas que apresentam respostas espectrais muito semelhantes entre si, e por esta razão não são adequadamente separáveis utilizando-se dados tradicionais em baixa dimensionalidade, constituindo-se portanto, em uma área adequada para testar a eficiência das técnicas propostas neste estudo. Com este objetivo, foram desenvolvidos uma série de experimentos utilizando as técnicas de regularização da matriz covariância. Nesta série de experimentos foram testados diferentes valores para o número das amostras de treinamento, para os valores numéricos de pares ( $\lambda, \gamma$ ), e diferentes números de bandas espectrais, isto é, dimensionalidade dos dados (de 5 a 185 bandas espectrais). Este conjunto de experimentos permitiu um teste bastante completo da eficiência do processo de regularização da matriz covariância, quando aplicado ao processo de classificação utilizando o classificador MVG, envolvendo classes espectralmente muito semelhantes. Algumas das principais conclusões obtidas, podem ser sintetizadas nas s figuras 1 e 2. Estas figuras ilustram comparativamente as exatidões obtidas no processo de classificação de duas das cinco classes que exibem um comportamento espectral muito semelhante (Corn-Min (milho plantio direto) e Soy-Clean (Soja plantio lavrado)). Para melhor ilustrar as vantagens oferecidas pelo processo de regularização, adotou-se um número baixo de amostras de treinamento (200 amostras por classe). Nestes dois experimentos foram adotados valores numéricos para  $\lambda$  e  $\gamma$  conforme explicitado nas figuras 1 e 2. O número de bandas espectrais (dimensionalidade dos dados) está colocado ao longo do eixo horizontal, e a exatidão do processo de classificação ao longo do eixo vertical. Os valores numéricos atribuídos aos parâmetros  $\lambda$  e  $\gamma$  encontram-se explicitados em ambas as figuras, para cada curva em particular. Pode-se observar claramente o efeito de Hughes na curva que representa a exatidão fornecida pelo classificador quadrático MVG. A exatidão da classificação tende inicialmente a aumentar na medida em que novas bandas espectrais são

incorporadas. A partir de um certo ponto, entretanto, o aumento da dimensionalidade dos dados tende a decrescer a exatidão. Como também seria de se esperar, o classificador linear, que para dados com baixa dimensionalidade fornece os valores mais baixos de exatidão, tende a melhorar a sua performance a medida em que a dimensionalidade dos dados aumenta. Também este fato vem a comprovar as suposições anteriormente expostas: Dado o menor número de parâmetros a serem estimados, este classificador tende a sofrer menos os efeitos negativos do aumento da dimensionalidade. As duas figuras ainda ilustram o fato de que o classificador que implementa a matriz covariância regularizada (*ADR*) tende a fornecer as acurácias mais altas, a medida em que a dimensionalidade dos dados aumenta.

**Figura 1**

**Figura 2**

## CONCLUSÕES

O presente estudo teve por finalidade verificar a eficiência de um particular método de regularização da matriz covariância no processo de classificação de imagens digitais em sensoriamento remoto, contendo classes espectralmente muito semelhantes. Experimentos foram realizados utilizando dados coletados pelo sensor AVIRIS. Os experimentos mostraram a eficiência da metodologia proposta, que faz uso de dois parâmetros ( $\lambda$  e  $\gamma$ ) na implementação de processo de regularização da matriz covariância. O processo mostrou ser capaz de atenuar sensivelmente o efeito de Hughes, o qual é claramente visível nos resultados obtidos a partir do classificador quadrático MVG. Os experimentos mostraram que, para os dados utilizados (imagem AVIRIS) o parâmetro mais relevante foi o parâmetro  $\lambda$ , que pondera as contribuições das matriz covariância estimada para cada classe individual com a matriz covariância comum estimada para todas as classes. Os experimentos colocaram ainda em evidência um fato com respeito ao comportamento do parâmetro  $\mathbf{g}$ : os melhores resultados em termos de acurácia no processo de classificação foram aqueles para os quais  $\mathbf{g} = 0$ . Em outras palavras, ao contrário dos resultados apresentados por Aeberhard (1994), os experimentos realizados neste estudo, sugerem que o parâmetro  $\mathbf{g}$  tem pouca relevância no processo de regularização. Uma possível explicação para este inesperado resultado consiste no seguinte: o parâmetro  $\mathbf{g}$  tende a contrabalançar a tendenciosidade na estimação dos autovalores da matriz covariância. Sempre que o número de amostras é pequeno, quando comparado com a dimensionalidade dos dados, os menores autovalores são estimados ainda menores, o oposto ocorrendo com os autovalores maiores. Em outras palavras, a aplicação de  $\mathbf{g}$  tende a tornar o hiperelipsóide Gaussiano menos elíptico, isto é, menos alongado e com uma dimensão um pouco maior na direção delgada. É um fato bem conhecido que em sensoriamento remoto, as classes estão freqüentemente distribuídas ao longo da direção de máxima dispersão dos dados. Incidentalmente, esta é a razão pela qual a transformada de Karhunen-Loeve (Componentes Principais), que é baseada na matriz covariância comum, funciona geralmente bem como um redutor de dimensionalidade. Em outras palavras, os autovetores associados aos maiores autovalores em cada classe são aproximadamente paralelos entre si. Nestas condições, e sendo os pontos centrais dos hiperelipsóides (vetor média) muito próximos entre si, o alargamento dos hiperelipsóides individuais tende a

umentar a região de superposição e portanto o erro de Bayes. Este fato, peculiar às classes de sensoriamento remoto envolvidas neste estudo pode ser uma explicação do comportamento do parâmetro  $\mathbf{g}$  no processo de regularização da matriz covariância neste particular estudo, diferentemente do comportamento deste parâmetro reportado por outros autores. Um problema relevante à metodologia avaliada neste estudo, consiste na identificação dos valores numéricos para os parâmetros envolvidos no processo de regularização (especialmente o parâmetro  $\lambda$ ). Efetivamente, não existe neste momento um procedimento definido que permita encontrar-se para cada caso em particular os valores numéricos mais adequados para estes parâmetros, que devem ser obtidos experimentalmente a partir das amostras de treinamento para cada classe. Deve-se notar que o custo computacional para encontrar as melhores configurações de  $\mathbf{I}$  e  $\mathbf{g}$  no processamento das amostras nos experimentos, vai crescendo à medida que aumenta a dimensionalidade.

Como sugestões para trabalhos futuros, pode-se citar: buscar um melhor entendimento com relação à contribuição do parâmetro  $\mathbf{g}$  para a separação de classes com comportamento espectral muito semelhantes usando dados reais em alta dimensionalidade de sensoriamento remoto (imagens hiperespectrais); estudar um método mais rápido para obtenção das melhores configurações de  $\mathbf{I}$ ,  $\mathbf{g}$  e a dimensionalidade, a fim de se aplicar o método *ADR*, com mais rapidez e eficiência.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Aeberhard, S., Coomans, D. And De Vel, O. Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings. **Pattern Recognition**, v.27, n.8, p. 1065-1077, 1994.
- Cornfield, J., Discriminant Functions, **Review of the International Statistical Institute**, v.35, p.142-153, 1967.
- Cortijo F.J., De La Blanca N. P., The Performance of Regularized Discriminant Analysis Versus Non-parametric Classifiers Applied to High-Dimensional Image Classification, **International Journal of Remote Sensing**, v.20, n.17, p.3345-3365, 1999.
- Friedman, J. H., Regularized Discrimination Analysis, **Journal of the American Statistical Association**, Theory and Methods, v.84, n.405, 1989.
- Fukunaga, K., Hayes R.R., Effects of Sample Size in Classifier Design, **IEEE Transaction Pattern Analysis and Machine Intelligence**, v.11, n.8, p. 873-885, Aug.1989.
- Fukunaga, K., **Introduction to Statistical Pattern Recognition**, 2<sup>a</sup> ed., School of Electrical Engineering, Purdue University, West Lafayette, Academic Press Inc. 1990.
- Haertel V., Landgrebe, D.A., On the Classification of Classes with Nearly equal Spectral Response in Remote Sensing Hyperspectral Image Data, **IEEE Transactions on Geoscience and Remote Sensing**, v.37, n.5, p.2374-2386, september 1999.
- Haertel V., Considerations about Classification of Remote Sensing Digital Image Data in High Dimensional Settings, **IX Simposio Latinoamericano de Percepción Remota**, Puerto Iguazú, Misiones, Argentina, november 2000.
- Hoffbeck J.P., Landgrebe D. A., Covariance Matrix Estimation and Classification With Limited Training Data, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.18, n.7, july 1996.

Hsieh, P., Landgrebe, D.A., Classification of High Dimensional Data, **TR-ECE 98-4**, School of Elec. And Comp. Eng., Purdue University, May 1998.

Hughes G. F. On the mean accuracy of statistical pattern recognizers, **IEEE – Transactions an Information Theory**, v.II-14, n.1, p.55-65, 1968.

Jain, A., Duin, R.P.W., Mao, J., Statistical Pattern Recognition: A Review, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.22, n.1, pg 4-37, 2000.

Jia, X., Richards, J., Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets, **IEEE Transactions on Geoscience and Remote Sensing**, v.32, n.2, p.274-280, march 1994.

Jia, X., Richards, J., Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, **IEEE Transactions on Geoscience and Remote Sensing**, v.37, n.1, p.538-542, January 1999.

Jimenez, L. O., Landgrebe,D.A., Hyperspectral data Analysis and Supervised Feature Reduction Via Projection Pursuit, **IEEE Transactions on Geoscience and Remote Sensing**, v.37, n.6, p. 2653-2667, november 1999.

Shahshahani, B. M., Landgrebe D. A., The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon, **IEEE Transaction on Geoscience and Remote Sensing**, v.32, n.5, p.1087-1095, september. 1994

Tadjudin, S., Landgrebe, D.A., Classification of High Dimensional Data with Limited Training Samples, **TR-ECE 98-8**, School of Elec. And Comp. Eng., Purdue University , May 1998.

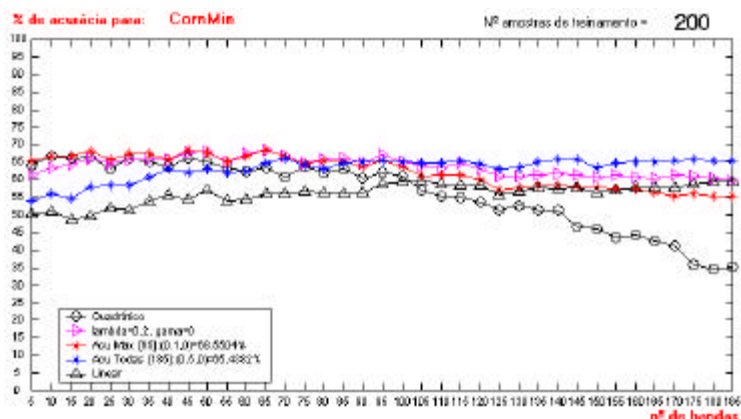


Fig. 1 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 200 amostras de treinamento

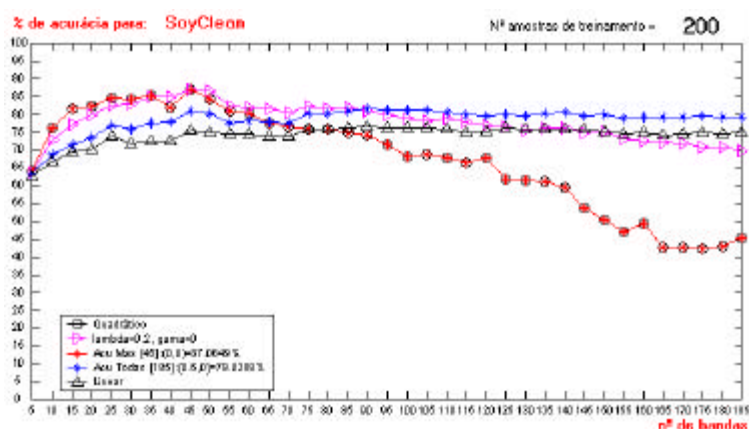


Fig. 2 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 200 amostras de treinamento