

## Estimação de Pesos Para Amostras Semi-Rotuladas em Classificadores Paramétricos

Atilio Efrain Bica Grondona<sup>1</sup>  
Vitor Francisco de Araújo Haertel<sup>1</sup>

<sup>1</sup> Universidade Federal do Rio Grande do Sul – UFRGS/CEPSRM  
Caixa Postal 15044 - 91501-970 - Porto Alegre - RS, Brasil  
{atilio.grondona, victor.haertel}@ufrgs.br

**Abstract.** In this paper it is investigated a methodology aimed at the estimation of the weights to be assigned to semi-labeled samples in the training of parametric classifiers when the number of training samples (pixels) are too small for the dimensionality of the data (bands), and the estimates of mean vector and the covariance matrix are poor. The proposed methodology includes two steps, the first step is only for estimate the weight of the semi-labeled samples, and the second step is for refine the weight estimated in the previous step. Initially it is investigated a procedure to estimate weights based on the classifier's decision function, where the semi-labeled samples are weighted according to probability of belongs to determined class. Next, these estimated are further refined by means of spatial context information, by use of the probabilistic relaxation to refine the weight of these semi-labeled samples. After this refinement, the mean vector and the covariance matrix are reestimated, and the parameters are used as input data in the parametric classifier. The methodology is tested by applying it to a set of hyperspectral image data. This data are from test site which had several classes spectrally similar, and is due this characteristic of the site which is propitious to work with hyperspectral sensors like AVIRIS.

**Palavras-chave:** classification, parameter estimation, spatial context, classificação, estimação dos parâmetros, contexto espacial.

## 1. Introdução

Dados em alta dimensionalidade, como por exemplo, aqueles disponibilizados por imagens hiperespectrais apresentam um grande potencial na classificação de imagens de cenas naturais. Classes espectralmente muito semelhantes, isto é, classes que apresentam estatísticas de primeira ordem (vetores de médias) muito similares podem ser separadas com um alto grau de acurácia em espaços de alta dimensionalidade, sob a condição das matrizes de covariância difiram significativamente entre si, conforme Fukunaga (1990) e Landgrebe (2003). Dados em alta dimensionalidade quando aplicados a classificadores paramétricos geram, entretanto, um número grande de parâmetros a serem estimados a partir das amostras de treinamento disponíveis (amostras rotuladas). Em situações reais, entretanto se dispõe em geral de um número apenas limitado de amostras para fins de treinamento e este fato resulta em estimativas pouco confiáveis para os parâmetros do classificador. Esta dificuldade afeta particularmente as estimativas para as matrizes de covariância, resultando no comportamento conhecido como fenômeno de Hughes. A acurácia produzida por um classificador paramétrico tende inicialmente a crescer na medida em que informação adicional, isto é, novas variáveis são acrescentadas aos dados. Em um determinado ponto, a acurácia atinge um máximo para em seguida passar a decrescer, na medida em que a dimensionalidade dos dados, isto é, o número de bandas espectrais, continua a crescer. Este comportamento reflete a dificuldade em se obter estimativas confiáveis para um número crescente de parâmetros.

Na literatura podem ser encontradas diferentes propostas visando minimizar este problema e desta forma ampliar o uso de imagens hiperespectrais na classificação de cenas naturais. Em termos gerais, as metodologias propostas podem ser agrupadas em três categorias: análise discriminante regularizada, descrita por Friedman (1989) e Aeberhard et al. (1984), redução na dimensionalidade dos dados por seleção ou extração de variáveis, conforme Fukunaga (1990) e Serpico et al. (2002), e o emprego de amostras semi-rotuladas, de acordo com Shashahani e Landgrebe (1994) e Jackson e Landgrebe (2001). Este estudo se insere nesta última abordagem. Com esta finalidade, amostras semi-rotuladas são selecionadas inicialmente e de uma forma iterativa durante o próprio processo de classificação, como proposto por Jackson e Landgrebe (2001) e Liczbinski e Haertel (2008). Neste estudo, o processo de estimação dos pesos para as amostras semi-rotuladas é investigado em maiores detalhes. Este processo produz uma estimativa inicial dos pesos que é em seguida refinada por meio de informações de contexto espacial. Na literatura se podem encontrar diferentes abordagens para fins de extração de informações de contexto espacial, como as propostas por Therrien (1989) e Solberg (1999). Uma das abordagens mais frequentemente empregada refere-se aos campos de Markov. Neste estudo se emprega um processo de relaxação. A metodologia investigada neste estudo é testada empregando-se dados hiperespectrais coletados pelo sistema sensor AVIRIS sobre a área teste denominada de *Indian Pine*, localizada no estado de Indiana, USA e o software necessário foi desenvolvido em ambiente MATLAB.

## 2. Metodologia de Trabalho

A etapa inicial consiste na seleção de um certo número de amostras semi-rotuladas e sua inclusão de uma forma ponderada na estimação dos parâmetros do classificador. Neste estudo segue-se uma abordagem similar à proposta em Jackson e Landgrebe (2001) e Liczbinski e Haertel (2008), propondo-se um critério mais elaborado na estimação dos pesos das amostras semi-rotuladas. O classificador aqui empregado é o classificador de Bayes, cuja expressão geral é dada por (Equação 1):

$$G_h(\mathbf{X}) = p(\mathbf{X} / \omega_h) P(\omega_h) \quad (1)$$

Sendo  $\mathbf{X}$  o pixel a ser rotulado,  $p(\mathbf{X} / \omega_h)$  a função densidade de probabilidade condicional à classe  $\omega_h$ , e  $P(\omega_h)$  a probabilidade *a priori* associada à mesma classe. Na suposição de uma distribuição Normal multivariada, a função de decisão (1) assume a forma (Equação 2):

$$G_h(\mathbf{X}) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Sigma_h|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_h)' \Sigma_h^{-1} (\mathbf{X} - \boldsymbol{\mu}_h)\right] P(\omega_h) \quad (2)$$

Sendo  $\boldsymbol{\mu}_h$  e  $\Sigma_h$  respectivamente o vetor de médias e a matriz de covariância associados à classe  $\omega_h$ , e  $p$  a dimensionalidade dos dados. O peso  $w_{h,i}$  associando o pixel  $\mathbf{X}_i$  à classe  $\omega_h$  pode ser estimado pela probabilidade *a posteriori*  $p(\omega_h / \mathbf{X}_i)$ , a qual pode ser calculada por meio do teorema de Bayes (Equação 3):

$$w_{i,k} = p(\omega_i / \mathbf{X}_k) = \frac{p(\mathbf{X}_k / \omega_i) P(\omega_i)}{\sum_{j=1}^m p(\mathbf{X}_k / \omega_j) P(\omega_j)} \quad (3)$$

e portanto (Equação 4):

$$w_{i,k} = \frac{G_i(\mathbf{X}_k)}{\sum_{j=1}^m G_j(\mathbf{X}_k)} \quad (4)$$

onde  $m$  representa o número de classes.

Na prática esta abordagem quando implementada diretamente pode apresentar dificuldades numéricas, devido aos valores muito pequenos resultantes da parte exponencial em (2), gerando estimativas não confiáveis para os pesos em (3). Este problema é comumente evitado empregando-se uma função monótona crescente de (2), mais especificamente a função logaritmo natural (Equação 5):

$$G_h(\mathbf{X}) = -\frac{1}{2} \ln |\Sigma_h| - \frac{1}{2} (\mathbf{X} - \boldsymbol{\mu}_h)' \Sigma_h^{-1} (\mathbf{X} - \boldsymbol{\mu}_h) + \ln P(\omega_h) \quad (5)$$

Embora numericamente mais adequada, esta alternativa apresenta a peculiaridade de que os valores resultantes resultam negativos. Para fins exclusivos de classificação dos pixels em uma imagem, este fato não apresenta nenhum inconveniente, pois a regra de classificação permanece inalterada: cada pixel é rotulado conforme a classe que produz o maior valor para a função de decisão (5). Como para todas as classes, os valores produzidos por (5) serão negativos, a classe vencedora será aquela que apresentar o menor valor absoluto. Esta peculiaridade inviabiliza, entretanto, o emprego direto da equação (3) para fins de estimação dos pesos. Neste estudo se propõe resolver este problema da seguinte forma. Inicialmente valores são calculados para cada pixel, em cada uma das  $m$  classes, aplicando a expressão logarítmica da função de decisão  $G(\mathbf{X})$

(5) em (4). Dado o fato anteriormente mencionado de que todos os valores de  $G(\mathbf{X})$  produzidos por (5) são negativos, os valores obtidos para os pesos  $w_{h,i}$  resultam ordenados inversamente. Neste estudo se propõe resolver este problema tomando-se o complemento destes valores (Equação 6):

$$w_{h,i}^* = 1 - w_{h,i} \tag{6}$$

Embora estes novos valores para os pesos estejam corretamente seqüenciados de acordo com os graus de pertinência de cada pixel a cada uma das  $m$  classes, torna-se ainda necessário efetuar a normalização, isto é, de que para qualquer pixel a soma dos pesos sobre as  $m$  classes deva ser igual à unidade (Equação 7):

$$w_{h,i}^{**} = \frac{\left( \frac{w_{h,i}^*}{\sum_{k=1}^m w_{k,i}^*} \right)}{m-1} \tag{7}$$

Um problema final a ser ainda considerado se refere ao valor assumido por cada pixel individual com relação à função densidade de probabilidade Normal multivariada (2). Observa-se pela expressão (4) que os pesos são estimados relativamente aos valores nas  $m$  funções de decisão  $G(\mathbf{X})$ . Nesta abordagem podem ocorrer situações em que o valor de  $G(\mathbf{X})$  para a classe vencedora seja reduzido, mas aqueles para as demais ( $m-1$ ) classes sejam ainda menores, resultando em um valor alto para o peso da classe vencedora. Torna-se, portanto necessário escalonar o peso da classe vencedora conforme sua posição efetiva na distribuição daquela classe, isto é, do valor de  $G(\mathbf{X})$  em termos do valor máximo desta função, o que ocorre para  $\mathbf{X} = \boldsymbol{\mu}$ . Na metodologia proposta neste estudo, se propõe multiplicar o peso da seguinte forma:

$$w_{h,j} = w_{h,j}^* \frac{G(\mathbf{X})}{G(\mathbf{M}_h)}$$

sendo  $\mathbf{M}_h$  o valor estimado para o vetor de médias  $\boldsymbol{\mu}_h$ , onde  $h$  representa a classe vencedora ( $\omega_h$ ). Então os parâmetros do classificador passam a ser estimados por um conjunto de amostras de treinamento, que inclui amostras rotuladas (peso unitário) e um determinado número de amostras semi-rotuladas, com peso estimado pelo procedimento acima descrito. As expressões empregadas para as estimativas amostrais do vetor de médias e da matriz de covariância ( $\mathbf{M}_h$ ,  $\mathbf{S}_h$ ) são dadas por Liczbinski e Haertel (2008) (Equação 8):

$$\mathbf{M}_h = \frac{1}{\sum_{k=1}^m w_{h,k}} \sum_{k=1}^{N_h} w_{h,k} \mathbf{X}_k \tag{8}$$

$$\mathbf{S}_h = \frac{1}{\left( \sum_{k=1}^m w_{h,k} \right) - 1} \sum_{k=1}^{N_h} w_{h,k} (\mathbf{X}_k - \mathbf{M}_h) (\mathbf{X}_k - \mathbf{M}_h)^t$$

onde  $N_h$  representa o número total de amostras de treinamento empregadas (rotuladas mais semi-rotuladas) para a classe  $\omega_h$ . Deve-se notar aqui, que os pesos referentes as amostras rotuladas são iguais à unidade.

A segunda etapa da metodologia consiste em refinar as estimativas para os pesos, por meio da introdução de informações de contexto espacial. Um dos métodos empregados com este objetivo consiste em métodos de relaxação probabilística, conforme os propostos por Therrien (1989) e Di Zenzo et al. (1987). Neste processo, a dependência espacial entre um pixel e seus vizinhos é estimada em um processo iterativo dentro de uma vizinhança previamente definida. Os pesos são ajustados de acordo com a expressão (Equação 9):

$$w_{i,h}^{t+1} = w_{i,h}^t \frac{1}{L} \left( 1 + (1/n) \sum_{j=1}^n \left( \sum_{k=1}^m r_{i,j,h,k} w_{j,k} \right) \right) \quad (9)$$

com os índices  $(t, t+1)$  indicando o número da iteração,  $r_{i,j,h,k}$  o coeficiente de compatibilidade,  $n$  o número de pixels na vizinhança,  $m$  o número de classes e  $L$  um fator de normalização, dado por (Equação 10):

$$L = \sum_{h=1}^m \left( 1 + (1/n) \sum_{j=1}^n \sum_{k=1}^m r_{i,j,h,k} w_{j,k}^t \right) \quad (10)$$

o coeficiente de compatibilidade  $r_{i,j,h,k}$  estima portanto a probabilidade de um pixel na posição central  $i$  de uma vizinhança pertencer à classe  $\omega_h$  ao mesmo tempo que seu vizinho na posição  $j$  pertence à classe  $\omega_k$  e é estimado por Di Zenzo et al. (1987) (Equação 11):

$$r_{i,j,h,k} = E \left[ \ln \frac{p(i \in \omega_h, j \in \omega_k)}{p(i \in \omega_h) p(j \in \omega_k)} \right] \quad (11)$$

uma estimativa amostral de (11) é dada por (Equação 12):

$$r_{i,j,h,k} = \ln \left( \frac{N(i, j, h, k) / N}{\left( \sum_{h=1}^m N(i, j, h, k) / N \right) \left( \sum_{k=1}^m N(i, j, h, k) / N \right)} \right) \quad (12)$$

onde  $N(i, j, h, k)$  representa o número de vezes em que um pixel na posição  $i$  pertence à classe  $\omega_h$ , ao mesmo tempo em que seu vizinho na posição  $j$  pertence à classe  $\omega_k$ , e  $N$  o número de pixels na imagem. O termo  $N(i, j, h, k) / N$  estima portanto a probabilidade conjunta  $p(i \in \omega_h, j \in \omega_k)$ , enquanto  $\left( \sum_{h=1}^m N(i, j, h, k) / N \right)$  e  $\left( \sum_{k=1}^m N(i, j, h, k) / N \right)$  estimam respectivamente as duas probabilidades marginais no denominador de (11). O processo de relaxação probabilística atualiza os valores estimados para os pesos  $w_{i,h}$  aplicando sucessivamente as equações (9), (10) e (12) em um procedimento iterativo, até que um critério de convergência seja satisfeito.

Deve-se aqui observar, que para fins de extração de informações de contexto espacial, é necessário levar em conta a existência de pixels que não pertencem a nenhuma das  $m$  classes em

consideração. Por esta razão, foi criada uma “classe de fundo”, composta por pixels cujo peso mais alto é menor do que um limiar previamente selecionado.

### 3. Resultados e Discussão

Vários experimentos foram realizados para fins de teste/validação da nova metodologia proposta. Cada experimento consiste na estimação da acurácia obtida no processo de classificação (acurácia do usuário e do produtor) variando-se a dimensionalidade dos dados. Para tornar os resultados comparáveis entre si, as demais variáveis (número de iterações, número de amostras rotuladas, número de amostras semi-rotuladas) foram mantidas constantes ao longo de cada experimento. Neste trabalho é apresentado e comentado os resultados de um, dos vários, experimentos realizados. As figuras abaixo ilustram os resultados gerados por este experimento, que tem como parâmetros: número de amostras de treinamento igual a 16 pixels, com o número de amostras semi-rotuladas igual a 30 pixels (Figura 1), 60 pixels (Figura 2) e 90 pixels (Figura 3). Em cada caso, a dimensionalidade dos dados variou dentro do intervalo [1, número de amostras rotuladas-1]. A Figura 4 ilustra os resultados pela primeira etapa da metodologia (sem a etapa de relaxação) e a Figura 5 com a inclusão da relaxação (informação de contexto espacial).

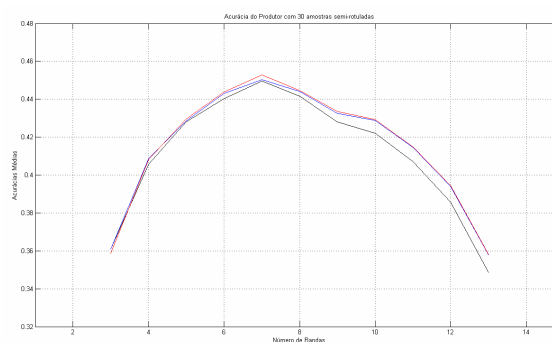
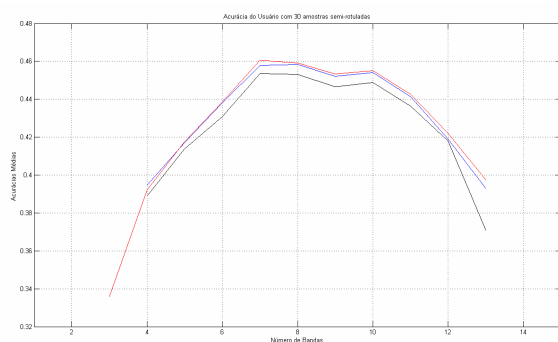


Figura 1. Acurácias médias, respectivamente, do usuário e do produtor. Em preto empregando somente as 16 amostras rotuladas, em azul com a adição de 30 amostras semi-rotuladas ponderadas, e em vermelho com o refinamento pela relaxação probabilística.

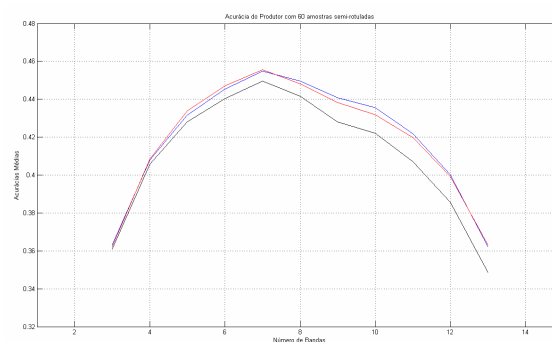
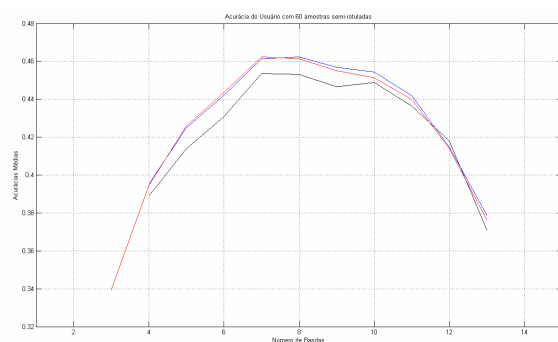


Figura 2. Acurácias médias, respectivamente, do usuário e do produtor. Em preto empregando somente as 16 amostras rotuladas, em azul com a adição de 60 amostras semi-rotuladas ponderadas, e em vermelho com o refinamento pela relaxação probabilística.

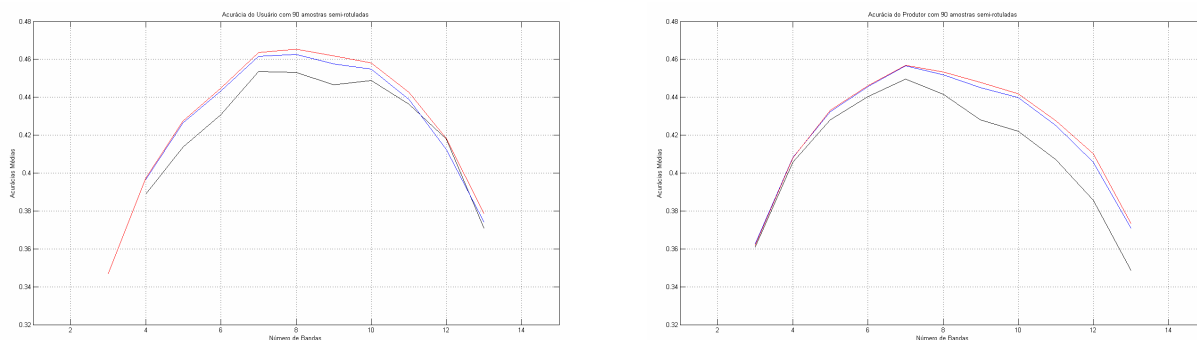


Figura 3. Acurácias médias, respectivamente, do usuário e do produtor. Em preto empregando somente as 16 amostras rotuladas, em azul com a adição de 90 amostras semi-rotuladas ponderadas, e em vermelho com o refinamento pela relaxação probabilística.

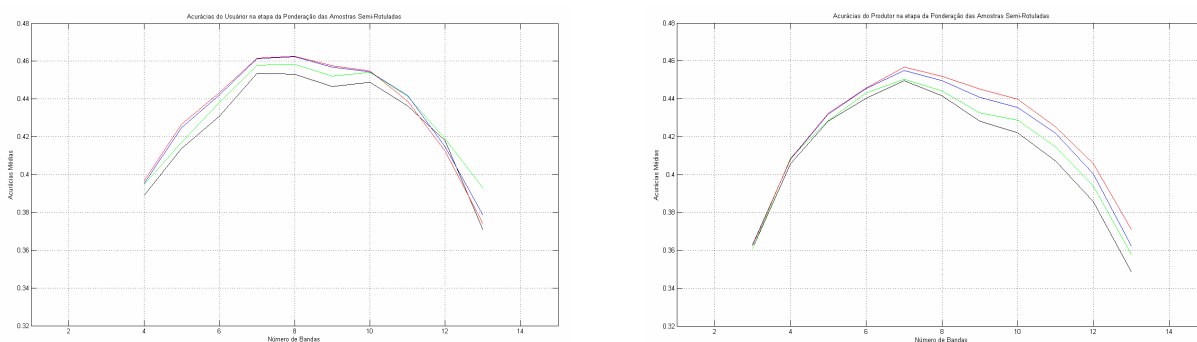


Figura 4. Acurácias médias, respectivamente, do usuário e do produtor na etapa da ponderação das amostras semi-rotuladas. Em preto somente com as 16 amostras rotuladas, em verde com a inclusão de 30 amostras semi-rotuladas, azul com a inclusão de 60 amostras semi-rotuladas ponderadas, e em vermelho a inclusão de 90 amostras semi-rotuladas ponderadas.

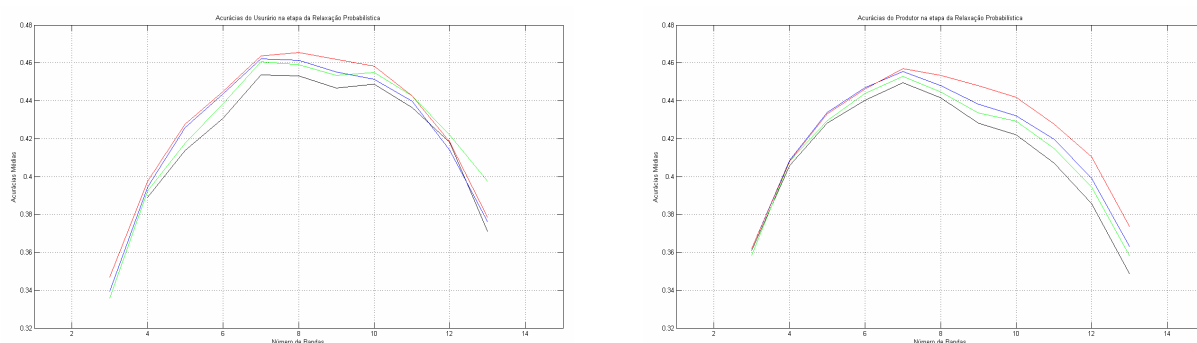


Figura 5. Acurácias médias, respectivamente, do usuário e do produtor na etapa da relaxação probabilística. Em preto somente com as 16 amostras rotuladas, em verde com a inclusão de 30 amostras semi-rotuladas, azul com a inclusão de 60 amostras semi-rotuladas ponderadas, e em vermelho a inclusão de 90 amostras semi-rotuladas ponderadas.

#### 4. Conclusões

Os experimentos acima descritos mostraram a eficácia de ambas as etapas da metodologia proposta. Isso pode ser visto com o aumento significativo da acurácia no pico do fenômeno de

Hughes. O aumento desta acurácia se dá não só no conjunto do processo, primeira e segunda etapas, como também, individualmente em cada etapa. Observa-se ainda que este aumento ocorre ao longo de todo o intervalo de variação na dimensionalidade dos dados.

### Referências Bibliográficas

Fukunaga, K. **Introduction to Statistical Pattern Recognition**. Boston: USA, 1990. 591 p.

Landgrebe, D. **Signal Theory Methods in Multispectral Remote Sensing**. Hoboken: USA, 2003. 508 p.

Friedman, J. H. Regularized Discriminant Analysis. **Journal of the American Statistical Association**, v. 84, n. 405, p. 165-175, 1989.

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.

Serpico, S. B., D'Inca, M., Melgani, F., and Moser, G., A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data. In: **Proceedings of SPIE - The International Society for Optical Engineering**, 2002. Agia Pelagia, Greece. v. 4885, p. 347-358.

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

Jackson, Q., and Landgrebe, D. An Adaptive Classifier Design for High-Dimensional Data Analysis With a Limited Training Data Set, **IEEE Transactions on Geoscience and Remote Sensing**, v. 39, n. 12, p. 2664-2679, 2001.

Liczbinski, C., and Haertel, V., A New Approach to Estimate A priori Probabilities in Remote Sensing Digital Image Classification, **Canadian Journal of Remote Sensing**, v. 34, n. 2, p. 135-142, 2008.

Indian Pine data disponível em:

<ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C>. Acesso em 2005.

Therrien, C. W., **Decision Estimation and Classification, An Introduction to Pattern Recognition and Related Topics**, New York: USA, 1989. 251 p.

Di Zeno, S., Bernstein, R., DeGloria, S., Kolsky, H. G., Gaussian Maximum Likelihood and Contextual Classification Algorithms for Multicrop Classification, **IEEE Transactions on Geoscience and Remote Sensing**, v. GE-25, n. 6, p. 805-814, 1987.

Byeungwoo, J., and D. Landgrebe, Classification with Spatio-Temporal Interpixel Class Dependency Contexts, **IEEE Transactions on Geoscience and Remote Sensing**, v. 30, n. 4, p. 663-672, 1992.

Jackson, Q., and Landgrebe, L., Adaptive Bayesian Contextual Classification Based on Markov Random Fields, **IEEE Transactions on Geoscience and Remote Sensing**, v. 14, n. 11, p. 2454-2463, 2002.

Jia, X., and Richards, J., Managing the Spectral-Spatial Mix in Context Classification Using Markov Random Field, **IEEE Geoscience and Remote Sensing Letters**, v. 5, n. 2, p. 311-314, 2008.

Solberg, A. H. S., Contextual Data Fusion Applied to Forest Map Revision, **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 3, p. 1234-1243, 1999.