

Um Algoritmo Geométrico para Geração de Funções Discriminantes

JOSÉ RICARDO PORTO SILVA

CLÉSIO LUIS TOZZI

Departamento de Engenharia de Computação e Automação Industrial

Faculdade de Engenharia Elétrica e de Computação

Caixa Postal 6101, UNICAMP

CEP 13081-970

Campinas, SP, Brasil

{jose,clesio}@dca.fee.unicamp.br

Abstract. Cluster analysis is a classification method which divides the pattern space into regions based on similarity measure of the feature vectors. It is used for nonsupervised learning to further classification of unknown elements. There exist many clustering algorithms; in general they have limitations about clusters shape or demand the pre-definition of convergence parameters. This work presents a technique to quickly determine piecewise linear discriminant functions using geometric characteristics of subsets on the pattern space. Examples with synthetic and real data are included.

Keywords: Cluster Analysis, Piecewise Linear Discriminant Functions, Pattern Recognition.

1 Introdução

No processo de reconhecimento de objetos, a aplicação de descritores quantitativos (isto é, descritores que atribuem um ou mais valores numéricos às suas funções de entrada) gera uma representação de cada objeto por um vetor de características. O conjunto de todos os possíveis vetores de características forma um espaço (de dimensão igual ao número de elementos do vetor) que é chamado de espaço de padrões ou espaço de características, onde cada vetor assume a forma de um ponto.

Análise de "clusters" é um método de classificação que divide o espaço de padrões em regiões (separando as classes) a partir de alguma medida de similaridade entre os vetores de características. Os métodos de classificação podem ser do tipo paramétrico ou não paramétrico. Como exemplos para o primeiro caso podem ser citadas a função perda, função discriminante de Bayes, função de máxima probabilidade [Andrews (1972)] que utilizam informações estatísticas dos dados, sua distribuição ou densidade de probabilidade. Porém, a eficiência desses métodos é comprometida se essas informações estatísticas não forem conhecidas a priori ou forem mal estimadas.

No segundo caso, uma função de distância é considerada para obtenção de um valor numérico para a similaridade entre os vetores de características. As funções mais comuns para este caso são: a inércia, pertinência fuzzy, distância euclidiana, distância de Mahalanobis e coeficiente de Tanimoto [Andrews (1972)]. Estes métodos geram as classes agrupando os

elementos mais próximos entre si. Os algoritmos K-means, K-NN [Sonka et al. (1993)], ISODATA [Bow (1984)] e CNN [Hart (1968)] são exemplos da aplicação desse método. Esse procedimento apresenta limitações devido a tendência dessas medidas de distância serem radiais, tornando as classes convexas e aproximadamente esféricas. Mesmo no caso da distância de Mahalanobis, que usa a matriz de covariância dos dados, e a distância euclidiana ponderada pela direção, haverá falhas em alguns casos extremos como as classes inter-penetrantes da Figura 1. Além disso, esse método só é bem sucedido se as distâncias entre classes forem maiores que as intra-classes.

Também podem ser encontrados na literatura métodos não paramétricos que usam distribuições gaussianas para definir as classes, como por exemplo [Fukunaga e Olsen (1970)], que apresentam baixa confiabilidade quando a função densidade de probabilidade difere consideravelmente de uma distribuição gaussiana.

Tanto para os métodos paramétricos como não paramétricos, o que se procura é a determinação de funções discriminantes, que são superfícies de separação, ou superfícies de decisão, que determinam os limites entre duas ou mais classes. Essas superfícies têm dimensão $n - 1$, onde n é a dimensão do vetor de características. As limitações ao uso desse tipo de superfície são a dificuldade de determiná-las e o seu grau. A forma matemática mais simples de separação é dada por superfícies discriminantes lineares, definidas pela Eq. (1).

$$f(X) = \sum_{i=1}^n a_i x_i + a_0 \quad (1)$$

Embora esta forma seja bastante eficiente e utilizada, algumas classes podem não ser linearmente separáveis, ou seja, podem ser necessárias superfícies quadráticas, cúbicas, etc para sua separação. A utilização de funções discriminantes lineares por partes pode eliminar esta dificuldade, mas em compensação gera outras, tais como a determinação do número de partes e seus intervalos de validade.

Com o objetivo de contribuir para a solução da questão colocada, este trabalho apresenta um algoritmo rápido e automático para geração das funções discriminantes lineares por partes. Para demonstrar a validade do algoritmo proposto, exemplos com dados sintéticos e reais são apresentados e seus resultados avaliados.

Este trabalho está organizado em quatro seções incluindo esta introdução. Na seção dois é descrito o algoritmo proposto. Na seção três são apresentados resultados e na seção quatro as conclusões.

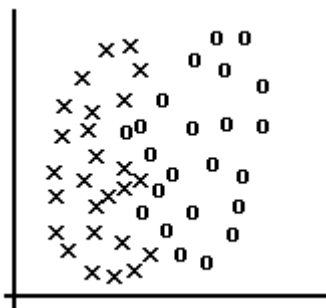


Figura 1: caso extremo com classes inter-penetrantes.

2 Descrição do algoritmo proposto

O fluxograma do algoritmo proposto é apresentado na Figura 2, e seu funcionamento é descrito nessa seção. Sendo um algoritmo que necessita de treinamento, este exige uma amostra na qual elementos de cada classe devem ser fornecidos com suas respectivas classificações (não necessariamente todas corretas).

Como primeiro passo do algoritmo são determinados os exemplares limitantes exteriores de cada classe segundo o método rotacional descrito na seção 2.1. Exemplares limitantes exteriores de cada classe são aqueles elementos que, quando ligados em ordem, formam o perímetro externo da distribuição espacial que o conjunto de amostras de cada classe apresenta no espaço de padrões.

No segundo passo do algoritmo é efetuada uma pré-classificação com base nos elementos limitantes (descrita com detalhes na seção 2.2). Se o resultado obtido não coincidir com o fornecido inicialmente, ações corretivas são tomadas segundo a Tabela 1, e volta-se ao primeiro passo enquanto não houver concordância entre os resultados.

No terceiro passo são calculadas as distâncias entre todos os elementos limitantes exteriores de cada par de grupos. Essa distância é arbitrada como infinita se um dos elementos for ocluso sob o ponto de vista do outro (o teste de oclusão é descrito na seção 2.3).

É selecionada a menor das distâncias entre elementos não interditados e calculado o ponto médio entre esses elementos. Os elementos utilizados são imediatamente definidos como interditados, para sua não reutilização. Esta interdição é realizada para não polarizar a superfície discriminante em torno desses elementos. Na sequência é verificado se existe alguma distância finita entre pares de elementos não interditados. Se houver, retorna-se ao passo anterior, caso contrário o algoritmo prossegue para o passo seguinte.

Os pontos médios obtidos são boas estimativas para a especificação das superfícies de separação pois, contendo exatamente os pontos que estão a igual distância entre cada duas classes, elas permitem igual variação para ambas as classes. Supondo que os exemplares fornecidos ao algoritmo determinam uma boa aproximação da função de distribuição das classes (seja ela de que tipo for), este método resulta numa estimativa quase ótima da função discriminante entre cada par de classes. Essa função é então gerada pela ligação entre os pontos médios obtidos, e será constituída por retas para um vetor de características bidimensional ($n=2$), planos para vetores tridimensionais ($n=3$), e hiperplanos para vetores de dimensão maior.

No passo seguinte é feita a verificação da validade dessa função discriminante utilizando o método especificado na seção 2.4. O teste deve ser realizado somente com os elementos limitantes do par de classes em questão, pois, por serem os mais exteriores, a função discriminante falhará para eles caso seja inválida. Se algum elemento limitante foi classificado erradamente, a função discriminante deve ser recalculada incluindo mais pontos médios para sua formação. Esse fato ocorre quando há poucos elementos em uma das classes, o que pode induzir a uma estimativa errada da superfície de separação. O ponto médio extra a ser usado é aquele entre o elemento mal classificado e seu vizinho mais próximo da classe oposta se ainda existe algum ponto médio relativo a esse elemento que não foi usado. Caso

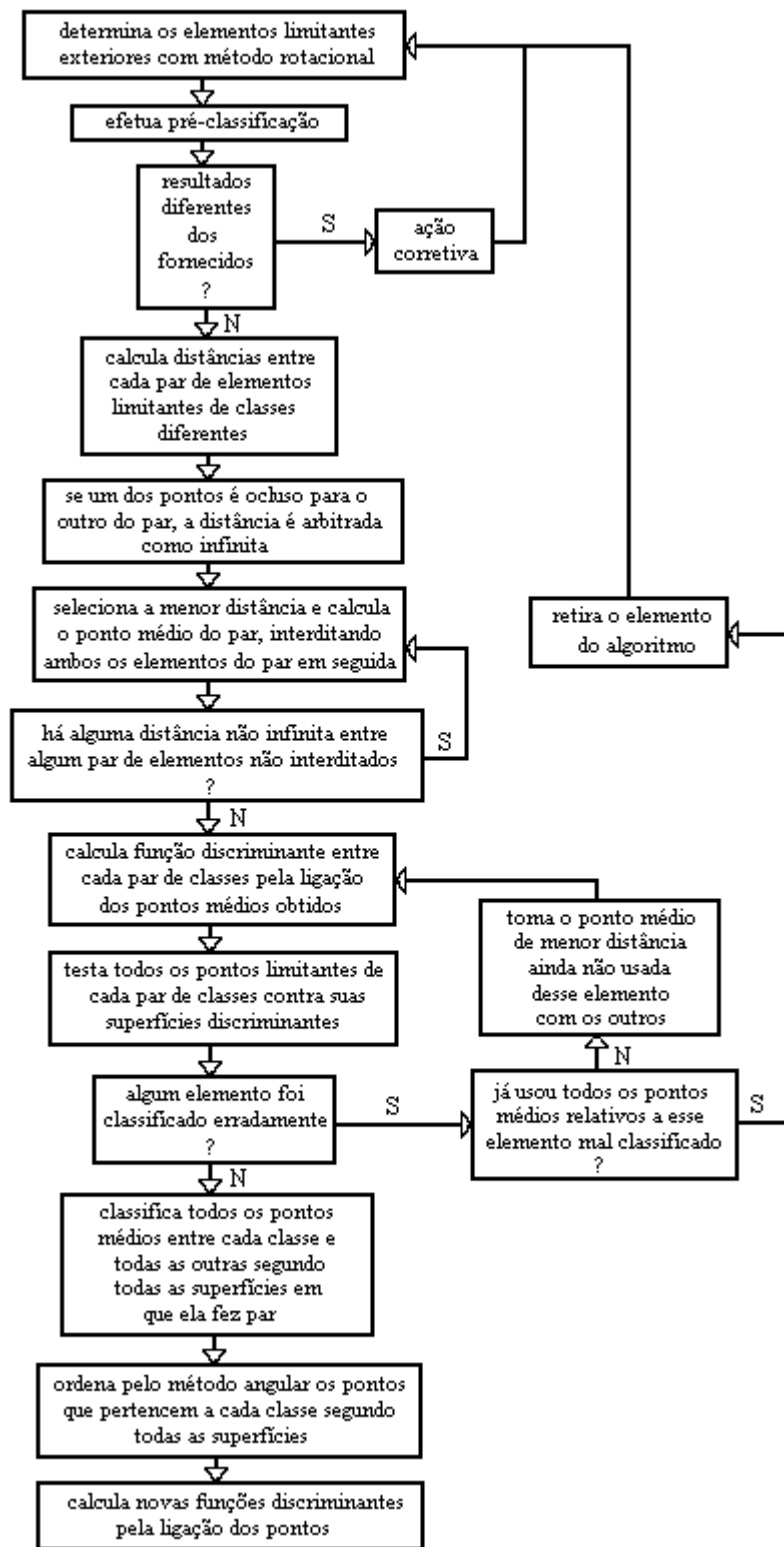


Figura 2: fluxograma do algoritmo proposto.

não exista, provavelmente a classificação fornecida para o elemento limitante é errada, ele deve ser eliminado da amostra e o algoritmo reiniciado.

No passo seguinte, para cada classe, todos os seus pontos médios, que formam as superfícies de decisão com todas as outras classes, são classificados segundo todas as superfícies discriminantes relativas à classe em consideração. São eliminados os pontos que não pertencem à classe segundo todas as superfícies discriminantes. Os pontos que não foram eliminados com certeza são limitantes da classe, portanto são usados para geração da função discriminante linear por partes. Para isso, eles são ordenados segundo sua posição angular em relação ao centróide da classe, sem alterar a ordem dos pontos que pertencem a uma mesma dicotomia.

Este método não garante que todos os pontos do espaço de padrões estarão contidos em alguma classe, isto é, pode haver regiões de indeterminação (ver Figura 6), entretanto os espaços interiores das classes são corretamente estabelecidos.

2.1 Método rotacional

Este passo do algoritmo seleciona e ordena os exemplares limitantes exteriores de cada classe. Na seqüência é descrita sua aplicação para o caso bidimensional, sendo possível sua extensão para espaços de maior dimensão. O processo toma como ponto inicial o elemento da classe mais distante do centróide. Todos os pontos da classe são girados (em torno do centróide) de um ângulo θ de forma a colocar esse ponto sobre a vertical que passa pelo centróide, e tornando sua ordenada maior que a do centróide. Dentre os pontos com abscissa maior que a do centróide, o de maior ordenada é selecionado como próximo elemento e colocado numa lista.

Este processo é repetido para cada novo ponto até voltar ao inicial. Pela descrição percebe-se que o algoritmo ordena os dados segundo o sentido horário, todavia, se houver uma fronteira de pequena concavidade, ao chegar ao ponto mais interior desta, a seleção avança para o vértice (final) dessa região. Como solução para esta restrição, o algoritmo é efetuado uma vez no sentido horário e outra no sentido anti-horário, e seus resultados são reordenados de forma a gerar a seqüência correta.

Outra restrição ocorre quando há fronteiras côncavas tão longas que o centróide situa-se fora do contorno limitante da classe, como se observa na imagem de uma lua crescente ou minguante. Nesse caso, o algoritmo como proposto passaria do início para o final da região côncava desconsiderando todos os

elementos intermediários. A ocorrência dessa restrição é identificada por $\theta \geq 90^\circ$, portanto, para esses valores, é usado um algoritmo alternativo para o método rotacional, chamado de método de varredura. Este algoritmo alternativo consiste em: para todo o elemento com ordenada maior que a do centróide e menor que a do último elemento processado pelo método rotacional, determinar a equação da reta que passa por esse elemento e o último selecionado. Para o caso em que a abscissa do centróide é maior que zero, é tomado como próximo elemento aquele cuja reta associada intercepta o eixo das ordenadas com menor valor. Caso contrário (abscissa negativa), é selecionado o elemento cuja intersecção ocorre com maior valor. Este processo pode ser visualizado imaginando-se um pêndulo fixo no último elemento processado, de comprimento inicial igual à distância entre este e o centróide, movendo-se para a esquerda até tocar algum outro elemento, daí originando a denominação "método de varredura".

2.2 Pré-classificação

Após a determinação dos elementos limitantes de cada classe, é gerado, no espaço bidimensional, um polígono com vértices nesses pontos, ou um hiperpolígono em espaços de maior dimensão. Na seqüência, o seguinte algoritmo é executado:

para cada elemento da amostra

para cada classe i

verificar a localização do ponto a que o elemento está associado no espaço de padrões. Se ele estiver dentro ou no perímetro do polígono definido pelos elementos limitantes da classe i então o elemento pertence à classe i .

Essa verificação é feita usando um algoritmo de ponto no interior de polígonos [Rogers (1988)] e outro de ponto sobre polígono.

Devido à forma de geração dos polígonos limitantes das classes, esse procedimento associa todos os elementos da amostra a alguma classe.

Se não houver coincidência entre a classificação obtida e a fornecida com a amostra é tomada uma das ações definidas na Tabela 1, senão prossegue o algoritmo calculando as distâncias entre elementos limitantes de classes diferentes.

As regras da Tabela 1 foram definidas empiricamente a partir de constatações realizadas para cada caso e são uma extensão das utilizadas no sistema S-MURICS proposto em [Ichino (1979)]. Através delas

são eliminadas as superposições e os elementos dúbios da amostra, aumentando a confiabilidade do algoritmo.

número de elementos da classe A classificados como sendo de B	número de elementos da classe B classificados como sendo de A	ação a ser tomada
1	0	retirar da amostra o elemento classificado erradamente da classe A
> 1	0	retirar da amostra o elemento da classe B mais próximo do centróide de A
1	1	retirar da amostra o elemento classificado erradamente que estiver mais distante do centróide da sua classe
> 1	1	retirar da amostra o elemento da classe B
> 1	> 1	retirar da amostra o elemento classificado erradamente que estiver mais próximo do centróide da outra classe

Tabela 1: ações corretivas para pré-classificação

2.3 Geração dos pontos médios

Para determinação da superfície discriminante entre cada par de classes são necessários apenas os elementos de fronteira não oclusos em relação a outra classe. Para isso é gerada uma matriz com as distâncias entre os elementos limitantes do par de classes consideradas, e traçada uma reta paramétrica entre os pares de pontos (com $0 \leq t \leq 1$). Sendo p_{ik} o i-ésimo elemento da classe k e p_{jm} o j-ésimo elemento da classe m, para o caso bidimensional:

$$\begin{aligned} y &= a_{ij}t + y(p_{ik}) \\ x &= b_{ij}t + x(p_{ik}) \end{aligned} \tag{2a}$$

com

$$\begin{aligned} a_{ij} &= y(p_{jm}) - y(p_{ik}) \\ b_{ij} &= x(p_{jm}) - x(p_{ik}) \end{aligned} \tag{2b}$$

onde $k \neq m$, e $y(p)$ e $x(p)$ representam a operação de extração da coordenada e abscissa do ponto p, respectivamente. A extensão para o caso geral é direta.

A seguir são calculadas as intersecções dessa reta paramétrica com os hiperplanos que contém as faces hiperpoligonais da distribuição espacial das duas classes. Se houver intersecção com o parâmetro t entre $0 < t < 1$ sobre os hiperpolígonos, um dos pontos é ocluso, e a distância entre os pontos é arbitrada como infinita.

Para o caso bidimensional (n=2), o teste pode ser realizado da seguinte forma: são calculadas as retas paramétricas que ligam os elementos de fronteira consecutivos de cada classe, isto é :

$$\begin{aligned} y &= c_{ik}s + y(p_{ik}) \\ x &= d_{ik}s + x(p_{ik}) \end{aligned} \tag{3a}$$

com

$$\begin{aligned} c_{ik} &= y(p_{(i+1)k}) - y(p_{ik}) \\ d_{ik} &= x(p_{(i+1)k}) - x(p_{ik}) \end{aligned} \tag{3b}$$

sendo p_{ik} o i-ésimo elemento da classe k, e são determinados os valores dos parâmetros t e s das intersecções das retas definidas pelas Eqs. (2) com as retas dadas pelas Eqs. (3) da classe k e da classe m. Se ocorrer $0 < t < 1$ e $0 < s < 1$ para qualquer das intersecções, então um dos pontos é ocluso e a distância entre eles é definida como infinita.

Após isso, as distâncias são ordenadas de forma crescente numa lista, desprezando aquelas entre elementos já usados em distâncias menores, e são calculados os pontos médios entre os elementos desta lista.

2.4 Classificação

A etapa de classificação é necessária no processo de determinação das superfícies discriminantes para eliminar os pontos médios que não pertencem às mesmas. Seja o caso de três classes circulares com desvio padrão igual a 0.1 e centróides sobre o eixo das abscissas, tendo coordenadas C1=(1,0), C2=(2,0) e C3=(3,0). A superfície discriminante da classe 1 deve passar pelo ponto médio entre os centróides das classes 1 e 2, portanto o ponto médio entre as classes 1 e 3 deve ser eliminado.

As superfícies discriminantes separam o espaço em duas regiões. A avaliação da sua função matemática para os pontos sobre a superfície resulta zero, e nos outros resulta em valores maiores ou menores que zero.

A partir do elemento da classe que gerou o ponto médio por onde a superfície discriminante passa, é avaliado o resultado da função e determinado o sinal (positivo ou negativo) atribuído para os elementos da

classe, e contra esse sinal devem ser comparados os resultados da avaliação de elementos desconhecidos.

No caso de uma superfície linear por partes deve ser determinado contra qual componente da superfície discriminante o elemento deve ser testado, ou seja, no intervalo de validade de qual das componentes o elemento desconhecido se situa. Para isso usa-se o conceito de superfícies de partição.

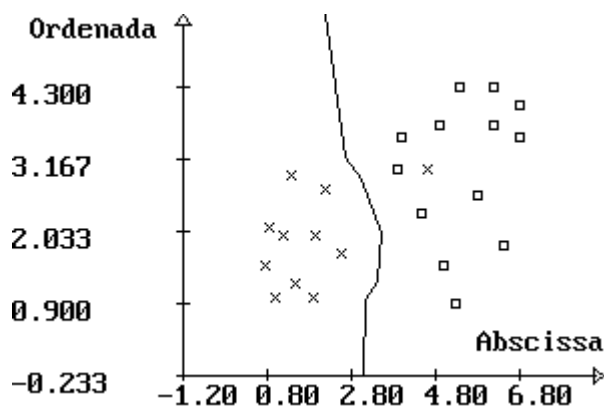


Figura 3: resultado obtido a partir de dados sintéticos.

As superfícies de partição têm a mesma dimensão das superfícies discriminantes e determinam a região de validade de cada uma das componentes. São definidas como a bissetriz do ângulo formado entre cada par consecutivo de superfícies de decisão, ou seja, para cada superfícies de decisão há duas de partição, exceto para a primeira e última superfícies de decisão (se essas não se unirem formando um "cluster" fechado), porque elas são ilimitadas. A primeira superfície de partição de uma superfície discriminante e a segunda da superfície discriminante anterior coincidem. As avaliações das duas superfícies de partição para um ponto resultam em valores não negativos se o ponto está no intervalo de validade da superfície discriminante correspondente. Dessa forma, um elemento desconhecido somente é testado contra a superfície de decisão cujas superfícies de partição o aceitam.

Se mais de um par de superfícies de partição aceita um novo elemento, e suas superfícies de decisão fornecem classificações contraditórias, então o elemento está ocluído por outras superfícies discriminantes. Nesse caso é necessário realizar o teste de oclusão para todas as superfícies de decisão cujas superfícies de partição aceitam o ponto, e rejeitar a classificação daquelas que apresentam oclusão. Isso é determinado calculando-se as equações paramétricas da reta que parte do elemento desconhecido e alcança o ponto médio da superfície discriminante sob teste com o valor do parâmetro igual a um, e verificando se há intersecção com outras

superfícies de decisão nos seus respectivos intervalos de validade com $0 < t < 1$. Isto é:

$$x_i = x_i(p) + a_{im}t \quad (4a)$$

com

$$a_{im} = \frac{\sum (x_i(pf_m) + x_i(pi_m))}{2(n-1)} - x_i(p) \quad (4b)$$

onde $x_i(pf_m)$ representa a coordenada i do ponto final da superfície de decisão m , e $x_i(pi_m)$ representa a coordenada i do ponto inicial da mesma superfície no caso da dimensão do vetor de características ser igual a 2. Para dimensão maior, os pontos iniciais e finais formam um hiperpolígono delimitante da superfície discriminante. Se m não for a primeira nem a última superfície de decisão é utilizada a equação (4b); se for, é utilizado um ponto qualquer da superfície (que esteja no interior do seu intervalo de validade) no lugar do ponto médio que é calculado na primeira parcela da equação (4b).

O valor do parâmetro para o ponto de intersecção da reta definida pelas Eqs. (4) com a superfície discriminante k é obtido de:

$$t = \frac{-a_{0k} - \sum_{i=1}^n a_{ik} x_i(p)}{\sum_{i=1}^n a_{ik} a_{im}} \quad (5)$$

com $1 \leq k \leq n^\circ$ de superfícies de decisão, e $k \neq m$.

Se houver intersecção da reta definida nas Eqs. (4) com a superfície k para valores de t entre $0 < t < 1$ e no intervalo de validade definido pelos pontos limitantes dessa superfície, então a classificação fornecida por m não é válida.

3 Resultados

O algoritmo descrito foi implementado em linguagem C para o caso bidimensional. Dados sintéticos e reais foram considerados na avaliação do seu desempenho. As Figuras 3 a 7 mostram resultados da separação entre classes. No caso da Figura 3 foi fornecida proposadamente errada a classificação de um elemento da amostra (há um X no interior da classe dos quadrados). Esse erro foi detectado pelo algoritmo, que eliminou esse elemento da amostra para efeito de determinação da superfície discriminante. Com isso, dentre os 24 elementos da amostra, somente este elemento foi classificado erradamente, ou seja, o índice de erros foi de 4,17%. Para os dados mostrados na

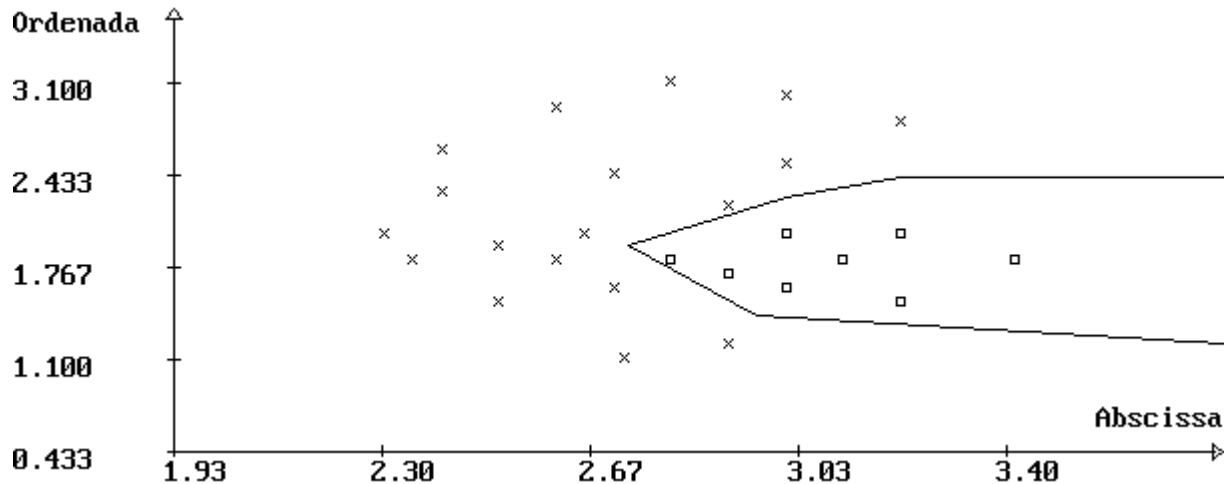


Figura 4: resultado obtido a partir de dados sintéticos formando um "cluster" convexo inserido num côncavo.

Figura 4 não houve erro de classificação. Na Figura 5 percebe-se um erro de classificação em 28 elementos, ou seja, um índice de 3,57%. Mas este elemento mal classificado está no interior do "cluster" da outra classe (como ocorre nas Figuras 3 e 6). Para os dados mostrados na Figura 6 houve dois erros em 35 amostras, tendo o algoritmo apresentado um índice de erro igual a 5,71%. Mas uma classificação intuitiva (proveniente da pergunta: dentro de qual "cluster" do espaço de padrões eles se situam ?) desses dois elementos coincide com a realizada pelo algoritmo. Para este caso ocorreu uma pequena região triangular de indeterminação. Na Figura 7 pode ser observada a classificação errada de um único elemento dentre 21, o que representa um percentual de erros igual a 4,76%. Esse elemento não foi corretamente identificado porque está dentro da região determinada pelos exemplares limitantes exteriores da classe dos quadrados, portanto foi retirado do processo de geração das funções discriminantes (segundo ação determinada pela Tab. 1) na etapa de pré-classificação.

Da análise realizada, foi observada para todos esses casos que a percentagem de erros de classificação foi baixa, demonstrando a eficiência do método proposto.

Apesar de seus vários passos, sub-algoritmos, uso de números em ponto flutuante com precisão dupla e de ter sido efetuado num computador lento para os padrões atuais (IBM PC 80386), a execução do algoritmo foi efetuada em menos de 1 segundo para todos os exemplos mostrados, o que possibilita inclusive o refinamento das superfícies discriminantes após a adição de cada novo elemento classificado.

4 Conclusões

Os resultados obtidos mostraram a validade, rapidez e eficiência do algoritmo proposto. Este algoritmo não

depende da seleção experimental de valores de limiar, como o apresentado em [Ichino (1979)], ou da pré-definição do número de componentes das funções discriminantes lineares por partes nem requer um longo treinamento para definição dos hiperplanos como o método proposto em [Foroutan e Sklansky (1987)]. Também não apresenta uma convergência lenta e dependente de parâmetros arbitrados como ocorre em [Spivak (1989)].

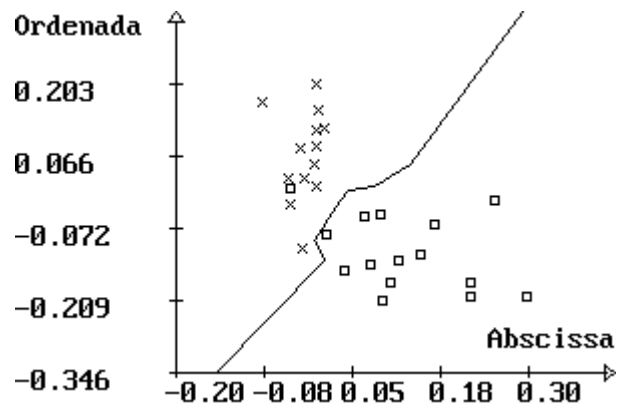


Figura 5: resultado obtido a partir de dados reais: descritores de Fourier de folíolos das espécies *Lonchocarpus campestris* (letras X) e *Lonchocarpus muehlbergianus* (quadrados).

Abordagem semelhante à apresentada neste trabalho é encontrada no caso de redes neurais, que também são usadas para classificação. Os perceptrons que compõem a rede podem ser representados por uma função discriminante linear seguida de uma binarização ou função sigmóide, portanto pode-se obter as superfícies de decisão a partir de redes neurais, como demonstrado em [Bobrowski (1991)]. Entretanto sua convergência é normalmente lenta e dependente do

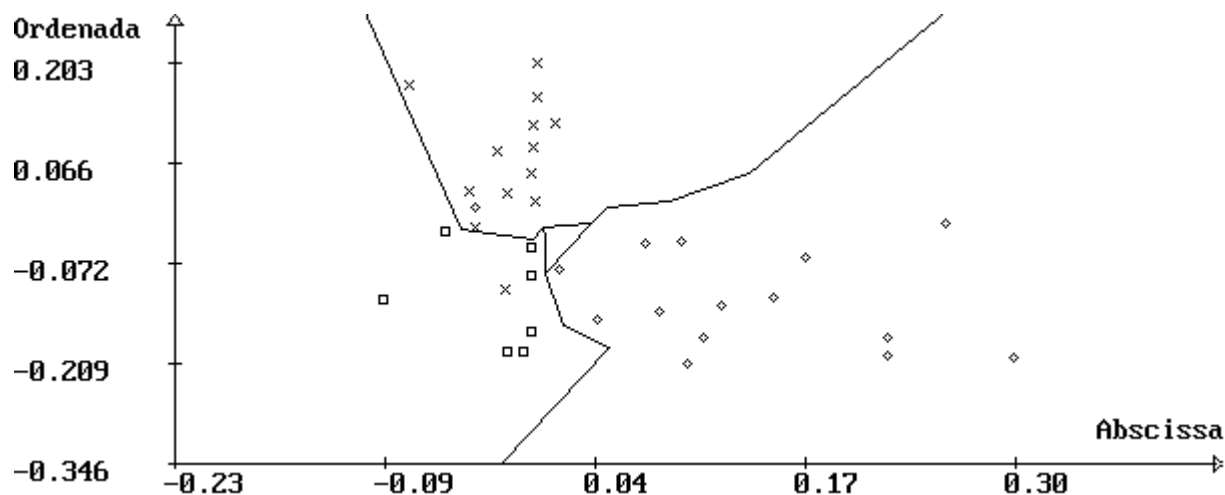


Figura 6: resultado obtido a partir de dados reais: descritores de Fourier de folíolos das espécies *Lonchocarpus campestris* (letras X), *Lonchocarpus guilleminianus* (quadrados) e *Lonchocarpus muehlbergianus* (losangos).

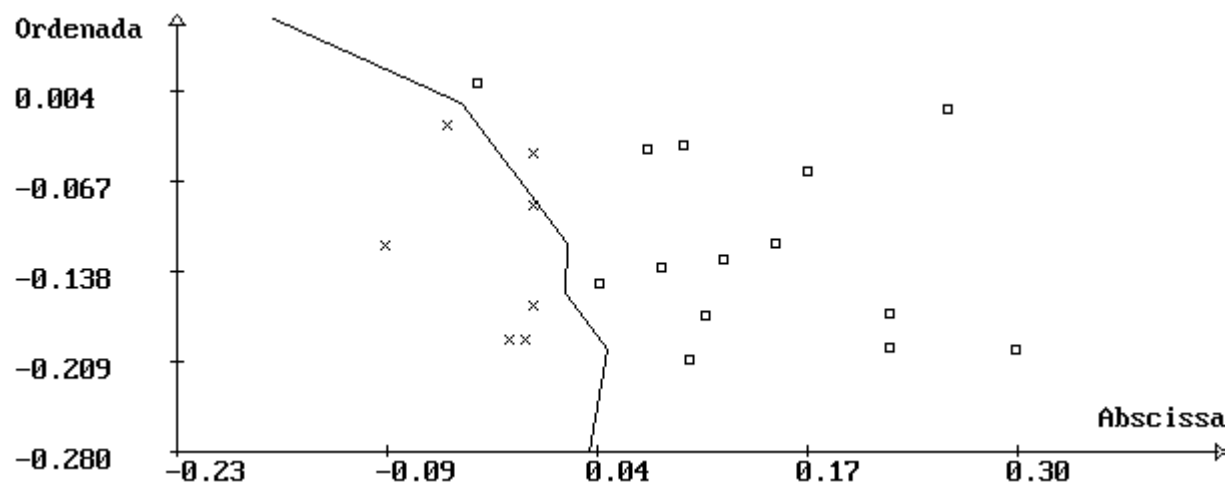


Figura 7: resultado obtido a partir de dados reais: descritores de Fourier de folíolos das espécies *Lonchocarpus guilleminianus* (letras X) e *Lonchocarpus muehlbergianus* (quadrados).

número de neurônios e arquitetura utilizados.

Referências

- H. C. Andrews, *Introduction to Mathematical Techniques in Pattern Recognition*, Wiley-Interscience, 1972.
- L. BoBrowski, "Design of Piecewise Linear Classifiers from Formal Neurons by a Basis of Exchange Technique", *Pattern Recognition* 24 (1991), 863-870.
- S. Bow, *Pattern Recognition*, Marcel Dekker Inc., 1984.
- I. Foroutan, J. Sklansky, "Feature Selection for Automatic Classification of Non-Gaussian Data", *IEEE Trans. Systems, Man, and Cybernetics* 17 (março/abril 1987), 187-198.

- K. Fukunaga, D. R. Olsen, "Piecewise Linear Discriminant Functions and Classification Errors for Multiclass Problems", *IEEE Trans. Inform. Theory* 16 (janeiro 1970), 99-100.
- P. E. Hart, "The Condensed Nearest Neighbor Rule", *IEEE Trans. Inform. Theory* 14 (maio 1968), 515-516.
- M. Ichino, "A Nonparametric multiclass Pattern Classifier", *IEEE Trans. Systems, Man, and Cybernetics* 9 (junho 1979), 345-352.
- D. F. Rogers, *Procedural Elements for Computer Graphics*, McGraw-Hill, 1988.
- M. Sonka, V. Hlavac, R. Boyle, *Image Processing, Analysis and Machine Vision*, Chapman & Hall, 1993.
- S. Spivak, "A Multisurface Method for Pattern Classification", *Pattern Recognition* 22 (1989), 587-591.