

CLASSIFICADOR DE MÁXIMA MARGEM COM NORMA ARBITRÁRIA: FORMULAÇÃO, ALGORITMO E RESULTADOS

SAULO M. VILLELA¹, RAUL F. NETO², SAUL C. LEITE³, ADILSON E. XAVIER¹

¹Programa de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro
Caixa Postal 68.511, 21941-972 Rio de Janeiro, RJ, BRASIL

²Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora
36036-330 Juiz de Fora, MG, BRASIL

³Coordenação de Matemática Aplicada e Computacional, Laboratório Nacional de Computação Científica
25651-075 Petrópolis, RJ, BRASIL

E-mails: {saulomv, adilson}@cos.ufrj.br, raulfonseca.neto@ufjf.edu.br, lsaul@lncc.br

Resumo – Este trabalho tem como objetivo apresentar uma nova formulação para o problema de obtenção de um classificador de máxima margem com norma arbitrária. O trabalho apresenta, também, um novo algoritmo baseado na solução de sucessivos sistemas de inequações com o uso de uma versão modificada do algoritmo de treinamento do perceptron. A sua grande vantagem consiste na sua flexibilidade, que permite a utilização de diferentes valores de norma, ao contrário de algumas formulações especiais existentes que se restringem à utilização das normas L_1 e L_∞ . O algoritmo foi chamado de IMA_p pois se trata de um caso especial do algoritmo IMA desenvolvido para a obtenção de classificadores de máxima margem com o uso da norma euclidiana.

Palavras-chave – Classificadores Lineares, Máquinas de Vetores Suportes, Normas conjugadas, Perceptron

1 Introdução

Um dos principais problemas relacionados à teoria do aprendizado de máquinas consiste no aprendizado de uma função ou classificador capaz de discriminar um conjunto de pontos pertencentes a duas classes distintas, em um espaço real. Uma questão importante está associada à capacidade de generalização deste classificador.

Neste sentido, tornou-se primordial a descoberta dos classificadores lineares de máxima margem. O projeto de tais classificadores se baseia em um novo princípio denominado princípio da minimização do risco estrutural, proposto por Vapnik (1995). Em sua formulação original, este problema foi desenvolvido no contexto da norma euclidiana L_2 . Assim, torna-se interessante o desenvolvimento de novas formulações flexíveis para a solução deste tipo de problema que explorem valores diferentes de norma, bem como novos algoritmos de treinamento.

1.1 Problema de Classificação Binária e Classificadores Lineares

Seja um conjunto de dados Z de cardinalidade m , denominado conjunto de treinamento, composto de um conjunto de vetores x_i e de um conjunto de escalares y_i . Cada vetor, rotulado por um valor escalar, está inserido em um espaço de dimensão d , $x_i \in R^d$, chamado de espaço de entrada do problema, representando uma respectiva amostra ou exemplo. Considerando que o valor de cada escalar y_i representa a classe de cada vetor x_i , tem-se para problemas de classificação binária, $y_i \in \{-1, +1\}$ para $i = \{1, \dots, m\}$. Para problemas linearmente separáveis, um classifi-

cador linear será representado no espaço de entrada por um hiperplano, dado pela seguinte equação:

$$f(x) = \langle w, x \rangle + b, \quad (1)$$

onde w representa o vetor normal ao hiperplano e b o valor do viés.

É possível representar este classificador em uma forma mais geral tomando-se cada ponto do espaço de entrada em um espaço denominado espaço de características ou Φ -space. Neste sentido, pode-se considerar a integração do viés da equação em uma componente adicional do vetor w , adicionando, também, uma componente +1 no vetor representativo de cada ponto. Assim a equação geral tem a forma:

$$f(x) = \langle w, \Phi(x) \rangle \quad (2)$$

A resposta do classificador poderá ser obtida através da aplicação de uma função sinal φ ao valor do discriminante relacionado à equação do hiperplano, ou seja:

$$\varphi(f(x)) = +1 \text{ se } f(x) \geq 0 \text{ ou } -1 \text{ se } f(x) < 0 \quad (3)$$

1.2 Algoritmo Primal do Perceptron

O algoritmo desenvolvido por Rosenblatt (1958) pode ser utilizado para a determinação do vetor w em um número limitado de iterações. A quantidade de iterações está relacionada à quantidade de atualizações do vetor de pesos e, conseqüentemente, à quantidade de erros cometidos pelo algoritmo.

Para uma determinada amostra do conjunto de treinamento, ocorrerá uma classificação incorreta se:

$$y_i(\langle w, \Phi(x_i) \rangle) < 0 \quad (4)$$

Neste sentido, pode-se adotar como função de perda a quantidade de amostras classificadas incorretamente. Esta função, definida como a função de perda 0-1, é descrita como:

$$J(w) = \sum_i 1 \mid \{\varphi(f(x_i)) \neq y_i\} \text{ ou} \quad (5)$$

$$J(w) = \sum_i \text{Max} \{0, \varphi(-y_i(\langle w, \Phi(x_i) \rangle))\} \quad (6)$$

Entretanto, sendo esta função constante por partes e, portanto, não diferenciável, torna-se mais apropriado a utilização de uma nova função de perda, linear por partes, dada pela soma negativa de todos os valores funcionais, também chamados de valores de margens, das amostras classificadas incorretamente. Ou seja:

$$J(w) = \sum_i \text{Max} \{0, -y_i \langle w, \Phi(x_i) \rangle\}, \quad (7)$$

tornando possível a utilização do método do gradiente.

Portanto, caso o problema seja linearmente separável no Φ -space, para se determinar uma solução que minimize a função de perda em relação ao vetor w , é necessário avaliar o vetor gradiente considerando, somente, a ocorrência das amostras classificadas incorretamente. Este processo, aplicado individualmente a uma amostra, resulta na seguinte regra de correção:

$$w_{(t+1)} = w_{(t)} + \eta \cdot \Phi(x_k) \cdot y_k, \quad (8)$$

sendo η a taxa de aprendizado.

2 Formulação de Máxima Margem

2.1 Perceptron de Margem Fixa

Admitindo a inclusão no sistema de restrições do modelo perceptron de uma margem fixa, dada por um parâmetro γ , a solução do novo problema consiste na determinação de uma solução viável para o sistema de inequações lineares na forma:

$$y_i \langle w, \Phi(x_i) \rangle \geq \gamma \quad (9)$$

Entretanto, caso não seja possível limitar o valor da norma L_2 do vetor w o sistema de inequações, se linearmente separável, apresentará sempre uma solução viável considerando o crescimento da norma e, conseqüentemente, do valor do produto interno, para qualquer valor de margem γ . Para se resolver este problema é necessário estabelecer alguma forma de regularização no sentido de controlar ou de limitar o valor da norma do vetor w .

Leite e Fonseca (2007) propõem uma nova formulação para o modelo perceptron, no sentido de garantir que o conjunto de exemplos guarde uma distância mínima em relação ao hiperplano separador sem limitar diretamente o valor da norma do vetor w .

Para tanto, é considerada a restrição de que cada amostra deva possuir um valor de margem geométrica correspondente superior ou igual ao valor estabelecido como margem fixa, sendo o valor da margem geométrica definido como o valor da margem funcional da respectiva amostra dividido pelo valor da norma euclidiana do vetor w . Isto equivale à realização do produto interno do vetor $\Phi(x_i)$ pelo vetor unitário de direção w , representado por $w/\|w\|_2$. Assim, deve-se resolver o seguinte sistema de inequações não lineares para determinado valor de margem fixa, representado pelo parâmetro γ_f :

$$y_i \langle w, \Phi(x_i) \rangle / \|w\|_2 \geq \gamma_f \quad (10)$$

Em função desta modificação, torna-se necessário reescrever a função de perda do modelo de forma a possibilitar a obtenção de uma nova regra de correção. A nova função será equivalente à soma dos valores das respectivas margens geométricas dos exemplos que erram, considerando o desconto do valor da margem fixa. Ou seja:

$$J(w) = (\sum_i \text{Max} \{0, \gamma_f - y_i \langle w, \Phi(x_i) \rangle / \|w\|_2\}) \quad (11)$$

Portanto, ao contrário do algoritmo básico do perceptron, considera-se também como erro aqueles exemplos que, embora classificados corretamente, não estejam a uma distância mínima, no sentido geométrico, do hiperplano separador.

A solução do sistema de inequações pode ser considerada como aquela que minimiza a função de erro J . Neste sentido, tomando-se o gradiente da função em relação ao vetor w , tem-se a seguinte correção, caso ocorra um erro, $y_k \langle w, \Phi(x_k) \rangle < \gamma_f \|w\|_2$, aplicada a uma determinada amostra $(x_k, y_k) \in Z$:

$$w_{(t+1)} = w_{(t)} - \eta (\gamma_f w / \|w\|_2 - \Phi(x_k) y_k) \quad (12)$$

2.2 Maximização da Margem

Para uma norma L_2 , o problema de maximização da margem em classificação binária é equivalente à determinação da maior distância euclidiana entre dois pontos de classes contrárias em relação ao hiperplano separador. Neste sentido, para toda possível hipótese relacionada à existência de um classificador linear, deve-se resolver o seguinte problema de otimização:

$$\text{Max}_w \text{Min}_{x_k \in X} \{y_k f(x_k) / \|w\|_2\}, \quad (13)$$

sendo $f(x_k) = \langle w, \Phi(x_k) \rangle$ a equação que define o classificador.

Considerando o valor da margem geométrica mínima, para um hiperplano com parâmetro w , como:

$$\gamma_g = \text{Min}_{x_k \in X} \{y_k f(x_k) / \|w\|_2\} \quad (14)$$

A máxima margem, no sentido geométrico, relativo à escolha do melhor hiperplano, será definida por:

$$\gamma^* = \text{Max}_w \{\text{Min}_{x_k \in X} \{y_k f(x_k) / \|w\|_2\}\} \quad (15)$$

Pelo fato desta margem ser válida para todos os pontos, tem-se, para qualquer amostra (x_k, y_k) do conjunto de treinamento, a relação $y_k f(x_k) / \|w\|_2 \geq \gamma^* \geq \gamma_g$. Sendo assim, pode-se reescrever o problema relativo à maximização da margem na forma:

$$\text{Max}_w \gamma_g$$

Sujeito a

$$y_k f(x_k) \geq \|w\|_2 \cdot \gamma_g$$

No contexto das redes de regularização, Girosi (1998), é possível a obtenção de um classificador de máxima margem, através da minimização do risco regularizado no espaço de características, ou seja:

$$\text{Min}_f (\sum_i c(x_i, y_i, f(x_i))) + \lambda \cdot \phi(f(x)) \quad (17)$$

Considerando $\|w\|_2 \cdot \gamma_g = 1$, a função de perda individual de um classificador de máxima margem será dada por:

$$\text{Max} \{0, 1 - y_k f(x_k)\} \quad (18)$$

Portanto, para um funcional de regularização relacionado ao valor da norma L_2 das hipóteses candidatas, tem-se:

$$\text{Min}_w \{(\sum_k \text{Max} \{0, 1 - y_k f(x_k)\}) + \frac{1}{2} w \cdot w^T \quad (19)$$

O que se torna equivalente à formulação original do problema proposta por Vapnik (1995):

$$\begin{aligned} &\text{Min } \frac{1}{2} w \cdot w^T \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1 \end{aligned} \quad (20)$$

3 Algoritmo de Margem Incremental (IMA)

Uma nova formulação para o problema de maximização da margem foi proposta por Leite e Fonseca (2007) e desenvolvida a partir de duas constatações importantes. Primeiramente, observando o fato de que na obtenção da máxima margem, os pontos ou vetores suportes de classes contrárias se encontram a uma mesma distância do hiperplano separador. Ou seja, considerando as margens das classes de rótulos positivo e negativo, tem-se $\gamma^+ = \gamma^-$, onde:

$$\gamma^+ = \text{Min } y_k f(x_k), \text{ para todo } x_k \in X^+ \quad (21)$$

$$\gamma^- = \text{Min } y_k f(x_k), \text{ para todo } x_k \in X^- \quad (22)$$

Em segundo lugar, observando a possibilidade da obtenção de soluções de larga margem, em um número finito de correções, na solução do problema do perceptron de margem geométrica fixa, na forma:

$$y_k f(x_k) \geq \gamma_f \cdot \|w\|_2, \quad (23)$$

para valores de $\gamma_f < \gamma^*$.

Neste sentido, propõe-se a solução aproximada do problema de máxima margem, considerando a maximização explícita e direta da margem geométrica. Neste contexto, deve-se resolver o seguinte problema de otimização:

$$\begin{aligned} &\text{Max}_w \gamma_g \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq \gamma_g \cdot \|w\|_2, k = 1, \dots, m, \end{aligned} \quad (24)$$

Considerando o fato de que a regularização do vetor w está implícita na função de perda apresentada na seção 2.2, ocorre uma limitação no crescimento do valor da norma, impedindo que a mesma escape para valores muito altos. Assim, torna-se possível computar diretamente o valor da maior margem geométrica, a qual se aproxima suficientemente da margem ótima, no sentido de garantir a construção de um classificador de máxima margem.

3.1 Técnica de Solução

A técnica de solução desenvolvida consiste de uma estratégia de aprendizado incremental, através da qual são obtidas sucessivas soluções do problema do perceptron de margem geométrica fixa, para valores crescentes de margem. Este parâmetro inicia com o valor zero, equivalente a solução original do algoritmo perceptron, e tem seus valores incrementados de forma consistente, até aproximar-se do valor da margem máxima. Ou seja, para um conjunto de valores $\gamma_f \in [0, \gamma^*]$, sendo:

$$\gamma_f^{t+1} > \gamma_f^t, \quad (25)$$

para $t = 1, \dots, T - 1$, $\gamma_f^1 = 0$, $\gamma_f^T \approx \gamma^*$, soluciona-se, sucessivamente, o problema de inequações não lineares:

$$y_k f(x_k) \geq \gamma_f \cdot \|w\|_2, k = 1, \dots, m, \quad (26)$$

sendo cada solução equivalente à solução do problema do perceptron de margem geométrica fixa.

Para a atualização a cada iteração, do valor da margem fixa, adotam-se duas regras, baseadas em uma estratégia de balanceamento, que garantem a convergência para a solução de máxima margem:

Primeira regra: caso a solução do problema forneça as margens, negativa e positiva, diferentes, pode-se dizer que a solução obtida não caracteriza uma solução de máxima margem. Portanto, corrigimos o valor da margem fixa na forma:

$$\gamma_f^{t+1} = (\gamma^+ + \gamma^-)/2, \quad (27)$$

onde γ^+ e γ^- , definidas anteriormente, representam as margens de parada das duas classes.

Pode-se observar que, neste caso, haverá sempre a garantia de solução do novo problema, já que a nova margem fixa é sempre inferior a margem ótima, ou seja, $\gamma_f^{t+1} = (\gamma^+ + \gamma^-)/2 < \gamma^*$. Tal condição deriva do fato de que se as margens negativa e positiva são desiguais, então a margem total não é máxima, implicando em $(\gamma^+ + \gamma^-) < 2 \cdot \gamma^*$. Também se tem a garantia de convergência, já que a nova margem fixa obtida é superior à margem fixa anterior, ou seja: $\gamma_f^{t+1} > \gamma_f^t$. Tal condição deriva-se da constatação das seguintes relações de exclusividade:

$$\gamma^+ > \gamma_f^t \text{ e } \gamma^- \geq \gamma_f^t \text{ ou } \gamma^+ \geq \gamma_f^t \text{ e } \gamma^- > \gamma_f^t, \quad (28)$$

Segunda regra: caso a solução do problema forneça as margens, negativa e positiva, iguais, pode ser que a solução obtida seja uma solução de ótimo local. Portanto, torna-se necessário garantir um acréscimo no valor da nova margem fixa, na forma:

$$\gamma_f^{t+1} = \gamma_f^t + \text{Max} \{ \Delta, (\gamma^+ + \gamma^-)/2 - \gamma_f^t \}, \quad (29)$$

sendo Δ uma constante de incremento positiva.

3.2 Número Total de Correções

O limite no número total de correções pode ser estimado utilizando-se o limite obtido na prova de convergência do algoritmo perceptron de margem fixa, Leite e Fonseca (2007). Introduzindo o valor da margem relativa à última iteração, $\gamma_f^T < \gamma^*$, tem-se:

$$t \leq (R^2 - (\gamma_f^T)^2) / (\gamma_{w^*} - \gamma_f^T)^2 \quad (30)$$

Caso o algoritmo seja formulado no sentido de garantir uma α aproximação para o valor da margem tem-se o valor da margem fixa escrito na forma:

$$\gamma_f^T = (1 - \alpha) \cdot \gamma_{w^*} \quad (31)$$

Substituindo este valor na expressão anterior, tem-se o número limite de correções dado por:

$$t \leq (R^2 - ((1 - \alpha) \cdot \gamma_{w^*})^2) / (\gamma_{w^*} - (1 - \alpha) \cdot \gamma_{w^*})^2 \quad (32)$$

ou

$$t \leq (R^2 - ((1 - \alpha) \cdot \gamma_{w^*})^2) / (\alpha^2 \cdot \gamma_{w^*}^2) \quad (33)$$

4 Formulações com Norma Arbitrária

Uma importante flexibilidade, a exemplo do algoritmo ALMA_p, Gentile (2001), fornecida pelo algoritmo de margem incremental está no fato de poder-se trabalhar livremente com qualquer norma diferenciável relacionada ao vetor w . Para tanto, para diferentes valores de margens fixas, considera-se a solução de sucessivos problemas na forma:

$$y_k f(\Phi(x_k)) \geq \gamma_f \|w\|_q, k = 1, \dots, m, \quad (34)$$

no sentido de minimizar a norma L_q , dada por $\|w\|_q = (\sum_i |w_i|^q)^{1/q}$, do vetor w e de se estabelecer uma solução de margem L_p , baseando-se no fato de que as normas conjugadas p e q satisfazem a relação $1/p + 1/q = 1$.

4.1 Formulações Especiais

Existem dois casos específicos para valores de norma L_1 e norma L_∞ nos quais a solução do problema de máxima margem resulta na solução de um único problema de Programação Linear (PL).

Primeiramente, considera-se a minimização da norma L_1 do vetor w definindo um hiperplano separador com margem L_∞ . Ou seja, a distância computada dos pontos ao hiperplano separador é tal que maximiza o valor da maior componente do vetor normal. Esta variante é aconselhável se for necessário a obtenção de soluções mais esparsas em relação às componentes do vetor w , como no processo de seleção de características. Neste caso, é possível constatar que a solução ou hiperplano proveniente da formulação L_∞ se posiciona sempre quase perpendicular ao eixo da maior componente, tornando-se dependente desta coordenada.

Considerando o estabelecimento de uma margem funcional de valor unitário, ou seja:

$$\text{Min}_{x_k \in X} \{y_k f(x_k)\} = 1, \quad (35)$$

obtém-se o seguinte problema de otimização:

$$\begin{aligned} &\text{Min } \|w\|_1 \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1 \end{aligned} \quad (36)$$

Substituindo o valor da norma L_1 do vetor w , tem-se:

$$\begin{aligned} &\text{Min } \sum_i |w_i| \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1 \end{aligned} \quad (37)$$

Observando o fato de que $w = w^+ - w^-$ implica em $|w_i| = w_i^+ + w_i^-$, Kecman e Hadzic (2000) propõem a solução do problema acima na forma de um único problema de PL:

$$\begin{aligned} &\text{Min } \sum_i (w_i^+ + w_i^-) \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1, \\ &w_i^+ \geq 0, w_i^- \geq 0 \end{aligned} \quad (38)$$

sendo $f(x_k) = \langle w^+ - w^-, \Phi(x_k) \rangle$.

De outra forma, considera-se a minimização da norma L_∞ do vetor w , definindo um hiperplano separador com margem L_1 , ou seja, a distância dos pontos

ao hiperplano separador é tal que maximiza o somatório em módulo das componentes do vetor normal. Pedroso e Murata (2001) propõem uma formulação alternativa para a definição de um hiperplano separador com margem L_1 , que culmina na solução de um único problema de PL. Baseado no fato de que $\|w\|_\infty = \text{Max}_i |w_i|$, o problema de maximização da margem pode ser reformulado como:

$$\text{Max}_w \text{Min}_{x_k \in X} \{y_k f(x_k) / \text{Max}_i |w_i|\} \quad (39)$$

Limitando o valor da margem funcional, de modo que $\text{Min}_{x_k \in X} \{y_k f(x_k)\} = 1$, o problema de otimização se torna equivalente a um problema de PL:

$$\begin{aligned} &\text{Max}_w \{1 / \text{Max}_i |w_i|\} \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1 \end{aligned} \quad (40)$$

ou

$$\begin{aligned} &\text{Min } z \\ &\text{Sujeito a} \\ &y_k f(x_k) \geq 1 \\ &z \geq +w_i, z \geq -w_i \end{aligned} \quad (41)$$

Mesmo com a possibilidade de se resolver um único problema de PL, percebe-se que a condução do processo de otimização no espaço primal para as formulações L_1 e L_∞ torna complexa a solução de problemas de altíssima dimensão, já que a matriz de coeficientes do problema de PL possuirá posto relacionado à quantidade destes parâmetros. Também, estas formulações estão restritas à utilização das normas L_1 e L_∞ .

4.2 Algoritmo de Margem Incremental com Norma Arbitrária (IMA_p)

Neste caso, para a solução da formulação L_p com norma variável, opta-se pela solução direta do sistema de inequações:

$$y_k f(\Phi(x_k)) \geq \gamma_f \|w\|_q, k = 1, \dots, m \quad (42)$$

Para tanto, torna-se necessário definir a função de perda do modelo perceptron de margem fixa em sua forma geral, relacionada à existência de uma norma q conjugada, ou seja:

$$J(w) = (\sum_i \text{Max}\{0, \gamma_f - y_i \langle w, \Phi(x_i) \rangle\} / \|w\|_q) \quad (43)$$

Neste sentido, tomando-se a derivada da função em relação ao vetor w , tem-se a seguinte expressão que define o vetor gradiente local:

$$\nabla_w J(w) = ((\gamma_f \varphi(w_i) \cdot |w_i|^{q-1}) / \|w\|_q^{q-1}) - y_i \Phi(x_i) \quad (44)$$

Caso ocorra um erro, $y_i \langle w, \Phi(x_i) \rangle < \gamma_f \|w\|_q$, a seguinte regra de correção será aplicada a uma determinada amostra $(x_k, y_k) \in M$:

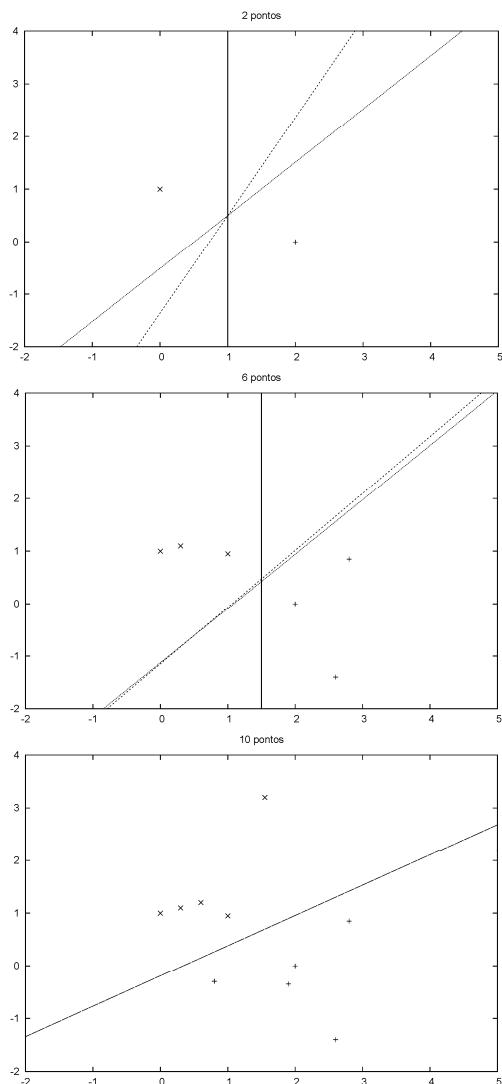
$$w_{(t+1)} = w_{(t)} + \eta ((\gamma_f \varphi(w_i) \cdot |w_i|^{q-1}) / \|w\|_q^{q-1} - y_i \Phi(x_i)) \quad (45)$$

Esta nova solução verifica-se como oportuna na medida em que podem ser utilizados quaisquer valores de norma variando entre L_1 e L_∞ preservando-se a mesma estrutura do algoritmo de margem incremental e modificando-se somente a equação de correção do vetor w .

5 Resultados

Inicialmente, para validação do algoritmo, foram testadas as formulações L_1 , L_2 e L_∞ (para ∞ foi utilizado o valor 100) nos conjuntos de pontos sugeridos por Pedroso e Murata (2001). A sugestão foi iniciar com os pontos (0,1) e (2,0), cada um pertencendo a uma classe diferente, e ir adicionando um ponto aleatório a cada classe. Uma classe é representada pelo símbolo \times e a outra pelo símbolo $+$.

Figura 1. Soluções de diferentes normas arbitrárias. A linha sólida corresponde à formulação L_∞ , a tracejada à L_2 e a pontilhada à L_1 .



Em seguida, utilizou-se o algoritmo IMA_p para as formulações L_1 , L_2 e L_∞ na solução de dois problemas de separabilidade linear e os resultados foram comparados com os obtidos pela solução do SVM-light. O primeiro, Mushroom, contém atributos de diversas variedades de cogumelos. A tarefa do classificador é determinar quais são comestíveis e quais não são. Possui 98 componentes e o conjunto foi dividido em 3763 amostras de treinamento e 1881 de teste. Apresentou 100% de acerto. O segundo, Sonar, consiste na classificação de sinais sonares. Conta com 60 componentes. O conjunto foi dividido em 169 amostras de treinamento e 39 de teste. Apresentou

67% de acerto. A esparsidade do vetor w foi medida em função do número de componentes com valor superior ao valor do percentual aplicado ao valor da maior componente.

Tabela 1. Comparação de valores para os algoritmos IMA_p ($p = 1, 2$ e ∞) e SVM-light aplicados aos problemas Mushroom e Sonar.

Algoritmo	Esparsidade			Margem	Acerto
	0,1%	1%	10%		
IMA_1	95	90	59	1,4644	100%
IMA_2	88	78	52	0,3646	100%
IMA_∞	13	9	9	0,0991	100%
SVM-light				0,3671	100%
IMA_1	60	60	56	0,0168	67%
IMA_2	60	57	48	0,0038	69%
IMA_∞	52	48	33	0,0007	67%
SVM-light				0,0043	67%

6 Conclusão

Nas simulações realizadas, pode-se constatar que os resultados alcançados foram iguais aos resultados analíticos obtidos pela solução das formulações especiais. Também, constatou-se a grande utilidade do algoritmo na obtenção de soluções esparsas, principalmente naqueles que possuem uma maior margem de separação. Como trabalhos futuros, sugere-se a utilização do algoritmo em problemas relacionados à seleção de características e, também, à análise de regressão, dada a existência de uma versão do algoritmo IMA aplicada ao problema de regressão, Fonseca, Borges e Leite (2008).

Referências Bibliográficas

Fonseca Neto, R., Borges, C. H. e Leite, S. C. (2008). An online training method based in support vectors for regression problems, *XXIX CILAMCE*, Macéio, AL.

Gentile, C. (2001). A new approximate maximal margin classification algorithm, *Journal of Machine Learning Research*, Vol. 2, 213-242.

Girosi, F. (1998). An equivalence between sparse approximation and support vector machines, *Neural Computation*, Vol. 10, no. 6, 1455-1480.

Kecman, V. e Hadzic, I. (2000). Support vectors selection by linear programming, *Proceedings of the International Conference on Neural Networks*, 5, 193-198.

Leite, S. C. e Fonseca Neto, R. (2007). Incremental Margin Algorithm, *Neurocomputing*, 71, 1550-1560.

Pedroso, J. P. e Murata, N. (2001). Support vector machines with different norms: motivation, formulations and results, *Pattern Recognition Letters*, 22, 1263-1272.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Review*, 65, 386-408.

Vapnik, V. N. (1995). The nature of statistical learning, *Springer*.