IP-LSSVM: A two-step sparse classifier
Autores: B.P.R. Carvalho, A.P. Braga

Apresentação: Paulo Sérgio Pereira Pessim

EEE928 - Técnicas Clássicas de Reconhecimento de Padrões

Programa de Pós-Graduação em Engenharia Elétrica - UFMG

Prof. Antônio Braga e Prof. Frederico Coelho

7 de Julho de 2021

Introdução

- A técnica SVM (do inglês, *Support Vector Machine*) é amplamente empregada.

Introdução

- A técnica SVM (do inglês, *Support Vector Machine*) é amplamente empregada.
  - Maximização de margem (Abordagem de Programação Quadrática).
  - Seleção dos Vetores de Suporte (Sensitividade dos Multiplicadores de Lagrange - Não nulos nas margens).

Introdução

- A técnica SVM (do inglês, *Support Vector Machine*) é amplamente empregada.
    - Maximização de margem (Abordagem de Programação Quadrática).
    - Seleção dos Vetores de Suporte (Sensitividade dos Multiplicadores de Lagrange - Não nulos nas margens).
- Desenvolvimento de Estratégias que demandam menos recursos computacionais.

Introdução

- A técnica SVM (do inglês, *Support Vector Machine*) é amplamente empregada.
  - Maximização de margem (Abordagem de Programação Quadrática).
  - Seleção dos Vetores de Suporte (Sensitividade dos Multiplicadores de Lagrange - Não nulos nas margens).
- Desenvolvimento de Estratégias que demandam menos recursos computacionais.
- *Least Squares Support Vector Machine - LSSVM*
  - Menos Recursos - Sistema de Equações Lineares.
  - Perda de *'sparseness'*.
  - Selecionar os Vetores de Suporte pelo critério *non-zero*.

Introdução

- A técnica SVM (do inglês, *Support Vector Machine*) é amplamente empregada.
  - Maximização de margem (Abordagem de Programação Quadrática).
  - Seleção dos Vetores de Suporte (Sensitividade dos Multiplicadores de Lagrange - Não nulos nas margens).
- Desenvolvimento de Estratégias que demandam menos recursos computacionais.
- *Least Squares Support Vector Machine - LSSVM*
  - Menos Recursos - Sistema de Equações Lineares.
  - Perda de *'sparseness'*.
  - Selecionar os Vetores de Suporte pelo critério *non-zero*.
- A relação entre Vetores de suporte $\times$ Custo computacional.
  - Melhorar a representação dos vetores de suporte na abordagem LSSVM.

Formulação - LSSVM

- Conjunto de Treinamento $\{x_i, y_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$ é mapeado por meio de funções de kernel.

FORMULAÇÃO - LSSVM

- Conjunto de Treinamento $\{x_i, y_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$ é mapeado por meio de funções de kernel.
- O mapeamento resulta em um problema linearmente separável na forma:

$$w^T \Phi(x) + b = 0$$

  - w - Vetor de Parâmetros,
  - b - Termo da variável de folga (*bias-term*)

Formulação - LSSVM

- Conjunto de Treinamento $\{x_i, y_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^n$, $y_i \in \{-1, +1\}$ é mapeado por meio de funções de kernel.

- O mapeamento resulta em um problema linearmente separável na forma:

$$w^T \Phi(x) + b = 0$$

  - w - Vetor de Parâmetros,
  - b - Termo da variável de folga (*bias-term*)

- Maximização de margem: Minimização da norma quadrática de w e do erro de treinamento

$$Mín_{w,b,e} \quad J_P(w,b,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i{}^2$$

sujeito a

$$y_i[w^T \varphi(x) + b] = 1 - e_i, \quad i = 1, \ldots N.$$

sendo γ um parâmetro de margem.

Formulação - LSSVM

- O Problema pode então ser reescrito como um sistema linear na forma $AX = B$, sendo

$$A = \begin{bmatrix} 0 & -Y^T \\ Y & H \end{bmatrix}, \quad X = \begin{bmatrix} b \\ \alpha \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

$$H = Z^T Z + \frac{I}{\gamma}, \quad Z = \begin{bmatrix} \varphi(x_1)y_1 & \dots & \varphi(x_1)y_1 \\ \vdots & \ddots & \vdots \\ \varphi(x_N)y_N & \dots & \varphi(x_N)y_N \end{bmatrix}.$$

- A saída do LSSVM pode ser computada por:

$$f(x) = sinal\left[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right], \quad K(x, x_i) = \varphi(x)\varphi(x_i).$$

- Os valores de α serão diferentes dos obtidos pela programação quadrática, uma vez que γ não impõe restrições como o parâmetro C, $(0 \leq \alpha \leq C)$.

Vetores de Suporte - LSSVM

- Seleção de acordo com $|\alpha|$.

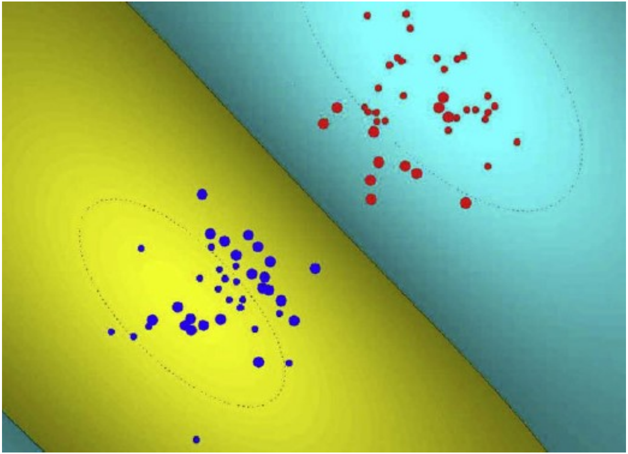

Figure: Vetores de Suporte LSSVM (Pontos maiores). Fonte: Artigo base

Relaçao de Margem - QP-SVM

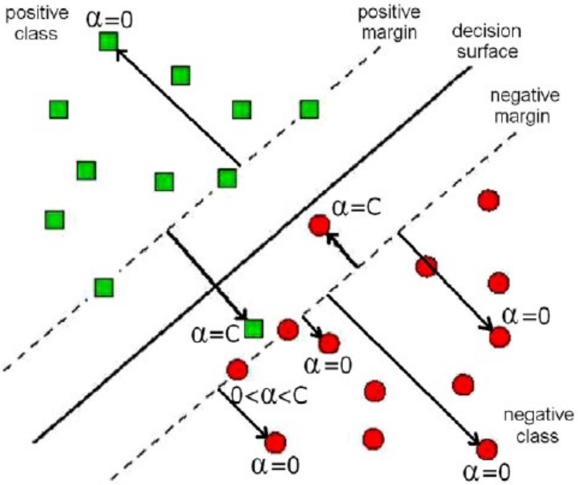

Figure: Valores de α e sua relação com a margem QP-SVM.

RELAÇAO DE MARGEM - LSSVM

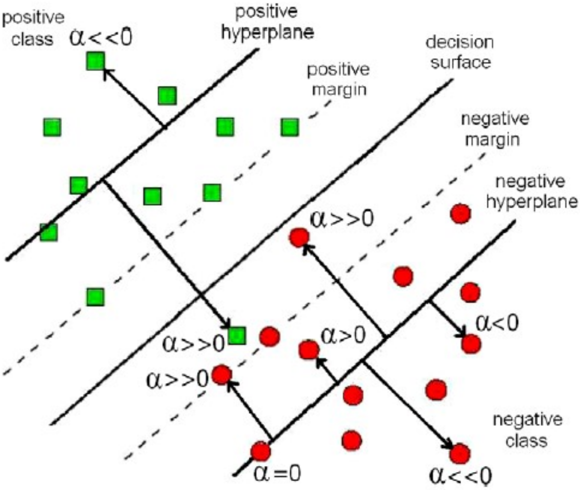

FIGURE: Valores de α e sua relação com a margem LSSVM.

NOVO CRITÉRIO DE RELEVÂNCIA

- x_i com $\alpha_i >> 0$ é um vetor de suporte: x_i localizado na borda entre as duas classes, ou então na região de classe oposta (Semelhante a α diferente de zero em QP-SVM).

NOVO CRITÉRIO DE RELEVÂNCIA

- x_i com $\alpha_i >> 0$ é um vetor de suporte: x_i localizado na borda entre as duas classes, ou então na região de classe oposta (Semelhante a α diferente de zero em QP-SVM).
- x_i com $\alpha \geq 0$ é eliminado: x_i é corretamente classificado e próximo ao hiperplano de suporte (Semelhante a região com $\alpha = 0$).

NOVO CRITÉRIO DE RELEVÂNCIA

- x_i com $\alpha_i >> 0$ é um vetor de suporte: x_i localizado na borda entre as duas classes, ou então na região de classe oposta (Semelhante a α diferente de zero em QP-SVM).
- x_i com $\alpha \geq 0$ é eliminado: x_i é corretamente classificado e próximo ao hiperplano de suporte (Semelhante a região com $\alpha = 0$).
- x_i com $\alpha < 0$ ou $\alpha_i << 0$ é eliminado: x_i é corretamente classificado e longe da superfície de decisão (Semelhante a região com $\alpha = 0$).

Vetores de Suporte Utilizando o Novo Critério

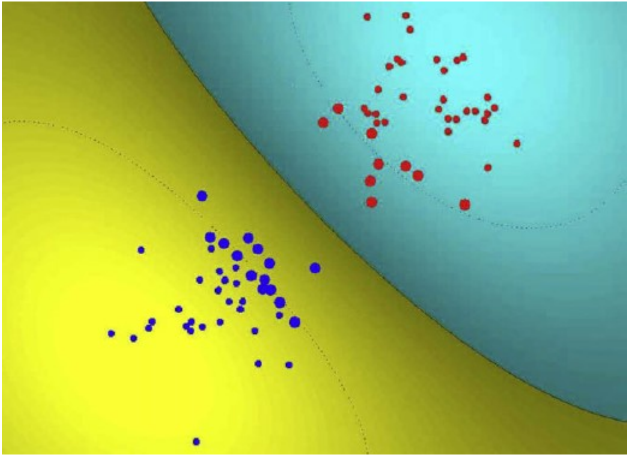

Figure: Vetores de Suporte para LSSVM utilizando o critério proposto. Fonte: Artigo base

Processo de Treinamento - IP-LSSVM

1. O sistema de equações lineares é resolvido com todos os vetores de treinamento, $X = A^{-1}B$.
2. $\tau \in \{0,1\}$ define a fração dos dados considerada como vetor de suporte.
3. Os vetores de treinamento são ordenados de acordo com seus valores de α.
4. A fração $(1-\tau)$ dos dados de treinamento com os menores valores de α é selecionada.
5. Uma matriz não quadrada A_2 é gerada removendo as colunas correspondentes aos dados selecionados no item anterior.
6. O novo sistema de equações é resolvido, $X_2 = {A_2}^{+}B$.
7. Os vetores de suporte são os pontos de treinamento de A_2.
8. Os valores de α e b são obtidos da solução de X_2.

Resultados

- O método proposto é comparado com outras técnicas presentes na literatura:
  - SVM (QP).
  - $LS^2 - SVM$.
  - Pruning.
  - ADA-PINV.
  - RRS+LS-SVM.
- São utilizados Kernels RBF e Linear.
- Quatro conjunto de dados do repositório da UCI são considerados, (Ionosphere, Pima Indians Diabetes, Bupa Liver Disorder e Tic Tac Toe).
- Dois conjuntos de dados sintéticos de duas dimensões são analisados para apresentar as superfícies obtidas.

RESULTADOS: SUPERFÍCIES

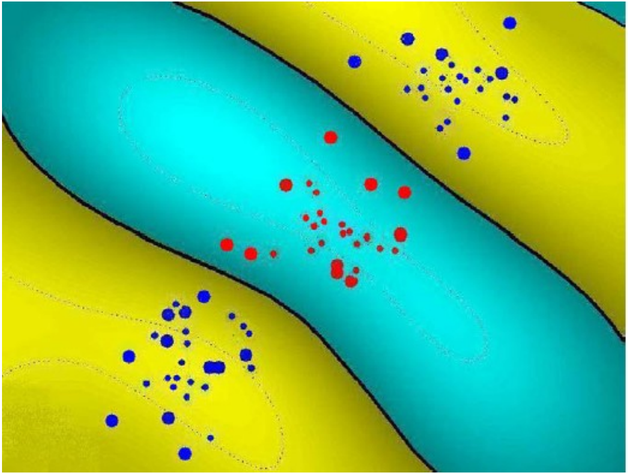

FIGURE: Superfície de Decisão obtida com o método $ADA - Pinv$ e kernel RBF (Problema dos três clusters). Fonte: Artigo base

Resultados: Superfícies

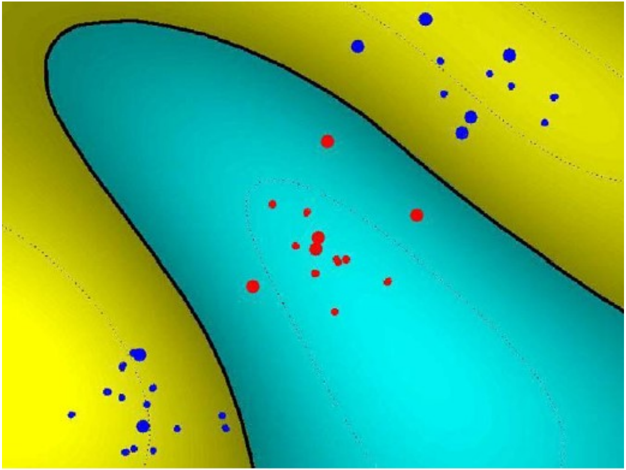

Figure: Superfície de Decisão obtida com o método Pruning e kernel RBF (Problema dos três clusters). Fonte: Artigo base

RESULTADOS: SUPERFÍCIES

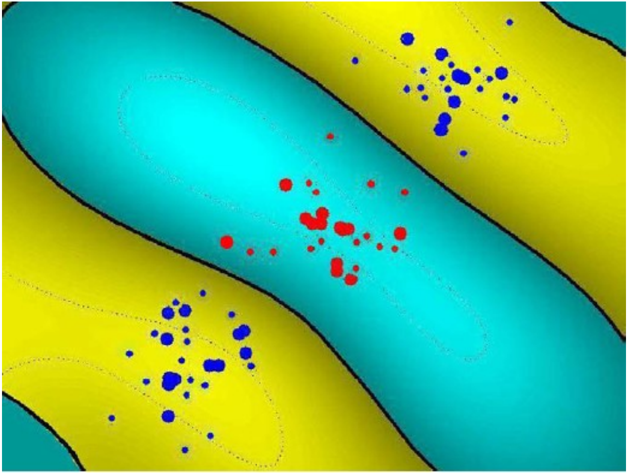

FIGURE: Superfície de Decisão obtida com o método $LS^2 - SVM$ e kernel RBF (Problema dos três clusters). Fonte: Artigo base

Resultados: Superfícies

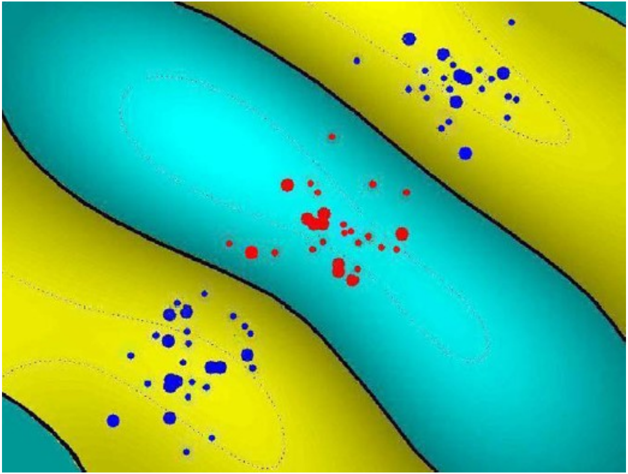

Figure: Superfície de Decisão obtida com o método $RRS + LS - SVM$ e kernel RBF (Problema dos três clusters). Fonte: Artigo base

Resultados: Superfícies

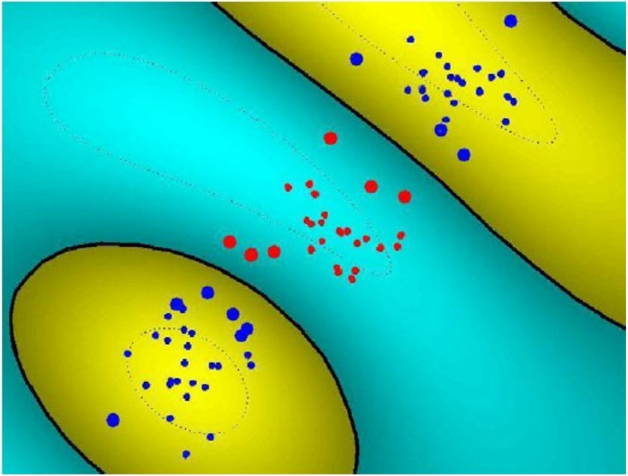

Figure: Superfície de Decisão obtida com o método *SVM* e kernel RBF (Problema dos três clusters). Fonte: Artigo base

RESULTADOS: SUPERFÍCIES

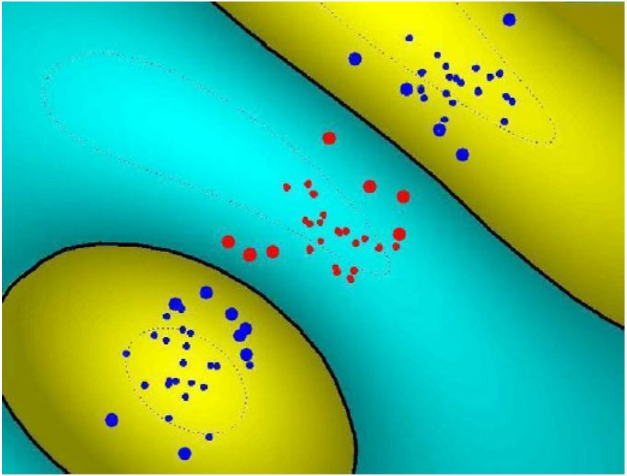

FIGURE: Superfície de Decisão obtida com o método $IP - LSSVM$ e kernel RBF (Problema dos três clusters). Fonte: Artigo base

RESULTADOS: ANÁLISE DO DESEMPENHO MÉDIO

- São apresentados os resultados médios para os 4 problemas considerando: Tempo de treinamento, Acurácia Média e Porcentagem de Vetores de Suporte ao longo de várias tabelas.
- Os resultados foram reduzidos a analise do método que obteve os melhores resultados.

RESULTADOS: ANÁLISE DO DESEMPENHO MÉDIO

- São apresentados os resultados médios para os 4 problemas considerando: Tempo de treinamento, Acurácia Média e Porcentagem de Vetores de Suporte ao longo de várias tabelas.
- Os resultados foram reduzidos a analise do método que obteve os melhores resultados.

TABLE: Métodos que apresentaram os melhores resultados médios com kernel linear

Base	Tempo de Treinamento	Acurácia Média	Vetores
Ionosphere	IP-LSSVM	IP-LSSVM/ SVM	IP-LSSVM
Pima Indians	Pruning	Pruning	RRS+LS-SVM
Bupa Liver	IP-LSSVM	IP-LSSVM	IP-LSSVM/ADA
Tic Tac Toe	Pruning	$LS^2 - SVM$	IP-LSSVM/ADA

- Nota-se o método proposto é o que mais aparece entre os melhores resultados.

Resultados: Análise do Desempenho médio

Table: Métodos que apresentaram os melhores resultados médios com kernel RBF

Base	Tempo de Treinamento	Acurácia Média	Vetores
Ionosphere	IP-LSSVM	IP-LSSVM/$*^1$	IP-LSSVM/$*^2$
Pima Indians	IP-LSSVM	SVM	LS^2-SVM
Bupa Liver	IP-LSSVM	LS^2-SVM	IP-LSSVM/ADA
Tic Tac Toe	IP-LSSVM	$ADA-PINV$	IP-LSSVM/ADA

$*^1 = LS^2 - SVM, \quad *^2 = LS^2 - SVM/ADA - PINV$

- Nota-se o método proposto é o que mais aparece entre os melhores resultados.
- Valores numéricos podem ser consultados no artigo.
- A diferença entre a acurácia média entre os métodos é sempre menor que 5%.

Conclusão

- Margem máxima é obtida pela formulação da programação quadrática, resultando em valores diferentes de zero para os multiplicadores de Lagrange próximo das margens de separação.

Conclusão

- Margem máxima é obtida pela formulação da programação quadrática, resultando em valores diferentes de zero para os multiplicadores de Lagrange próximo das margens de separação.
- Alternativas a programação quadrática (LSSVM).

Conclusão

- Margem máxima é obtida pela formulação da programação quadrática, resultando em valores diferentes de zero para os multiplicadores de Lagrange próximo das margens de separação.
- Alternativas a programação quadrática (LSSVM).
- As alternativas não são capazes de manter todas as características da formulação original.

Conclusão

- Margem máxima é obtida pela formulação da programação quadrática, resultando em valores diferentes de zero para os multiplicadores de Lagrange próximo das margens de separação.
- Alternativas a programação quadrática (LSSVM).
- As alternativas não são capazes de manter todas as características da formulação original.
- Em particular, perda de dispersão (*'sparseness'*) no vetor dos multiplicadores de Lagrange.

Conclusão

- Margem máxima é obtida pela formulação da programação quadrática, resultando em valores diferentes de zero para os multiplicadores de Lagrange próximo das margens de separação.
- Alternativas a programação quadrática (LSSVM).
- As alternativas não são capazes de manter todas as características da formulação original.
- Em particular, perda de dispersão (*'sparseness'*) no vetor dos multiplicadores de Lagrange.
- O método desenvolvido IP-LSSVM apresentou-se uma alternativa viável para detectar os vetores de suporte na margem, mantendo a simplicidade computacional do método LSSVM.

Agradecimentos

- A todos os presentes pela atenção e suporte

Email:
`paulopessim777@hotmail.com`