

Parametrização da SVM

Abordagem por Evolução Diferencial

Thiago Chakib Mattar Leydecker

Abstract—Um grande desafio em reconhecimento de padrões e problemas de classificação está na definição dos parâmetros do modelo. A temática deste trabalho consiste então na definição dos parâmetros de custo e largura da função de kernel de uma SVM, i.e. *Support Vector Machine*, a partir de uma abordagem de otimização que utiliza um algoritmo diferencial evolutivo. A comparação do método proposto é feita com a abordagem clássica de validação cruzada para várias bases de dados, que incluem um problema de classes desbalanceadas.

I. INTRODUÇÃO

Reconhecer faces, compreender um conjunto de palavras, ler símbolos manuscritos e identificar falhas e problemas são exemplos clássicos que podem ser abordados como sistemas de reconhecimento de padrões. Neste cenário, na atual era digital, é essencial compreender atividades que podem ser executadas por uma máquina e construir sistemas para fornecer apoio para tomadas de decisões e automação de processos com base apenas nos dados e informações disponíveis. A evolução da computação na era digital fornece informações e capacidade de processamento para tornar essencial a incorporação de técnicas de aprendizado estatístico em atividades como engenharia, medicina, direito, gestão empresarial e comercial. Existe generalização suficiente para incorporar um sistema de reconhecimento de padrões para melhorar e eficiência de praticamente qualquer atividade dos dias atuais.

Neste contexto, é importante definir o conjunto de técnicas e atividades que fazem parte do processo de reconhecimento de padrões. A partir nas informações disponíveis, o primeiro passo consiste no pré-processamento dos dados, que consiste na construção da significância de cada amostra ou característica associada ao problema. Em seguida, é importante extrair e definir as características relevantes apresentadas nos dados para então construir

um modelo de classificação que fornece a base para o processo decisório.

A etapa abordada neste trabalho consiste na parametrização do modelo de uma máquina de vetor de suporte. O primeiro parâmetro γ define a largura da função de kernel de base radial que realiza a transição dos dados para o espaço das verossimilhanças, de modo a tratar as não-linearidades pertencentes ao problema. O segundo parâmetro de custo pondera a função de custo interna ao modelo, que maximiza a largura da banda de vetores de suporte, e minimiza o erro quadrático da superfície de classificação. É importante ressaltar que a combinação ideal desses parâmetros é essencial para a construção de um modelo conciso, que maximiza a generalidade da estrutura de classificação, de modo a evitar *overfitting* aos dados de treinamento.

II. METODOLOGIA

É importante definir como o problema é tratado a partir de uma abordagem de otimização. Dado um conjunto de parâmetros $\{x_1, x_2, \dots, x_i\}$, e uma função de custo J , têm-se como objetivo encontrar o conjunto de valores dos parâmetros que minimizam a função J . Sendo assim, para o problema em questão, define-se como parâmetros $\{\gamma, C\}$ correspondentes à largura da função de kernel e ao custo, respectivamente. Além disso, define-se como função de custo $J = -AUC$, i.e. *Area under curve*, cuja medida expressa a combinação da especificidade e sensibilidade explicitadas na curva ROC e definem a eficiência da classificação, inclusive para classes desbalanceadas. Quanto maior a AUC, melhor a eficiência do modelo, e, para este caso, quanto menor J , melhor a eficiência do modelo.

A Figura 1 apresenta a superfície de busca da função de custo para a base de dados *Breast*

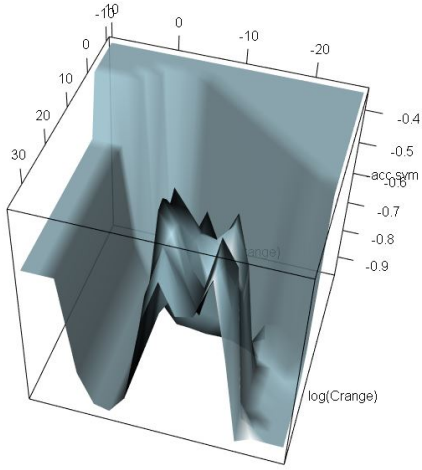


Fig. 1. Superfície da função de custo para a base Breast Cancer.

Cancer. É possível observar características não-lineares, multimodais e não-convexas. Além disso, a avaliação de J apresenta um fator estocástico associado com a variabilidade da amostragem dos dados de treinamento e alto de tempo de execução. Trata-se então de um problema de otimização de complexidade elevada que justifica a abordagem baseada em evolução diferencial. A comparação da eficiência do método é realizada com a abordagem clássica de validação cruzada.

A. Evolução Diferencial

Seja um problema de otimização não-linear irrestrito de variáveis contínuas, é possível resolvê-lo para encontrar o ponto de mínimo ou máximo de uma função por meio de um algoritmo de evolução diferencial (DE). Proposto por Reiner Storn e Kenneth Price em 1995, o algoritmo utiliza como mecanismo básico de busca o operador de mutação diferencial, e, por apresentar simplicidade de implementação, robustez e eficiência, auto adaptação e versatilidade, é muito popular na otimização não linear. O mecanismo clássico de busca do DE utiliza vetores-diferença, a partir de dois indivíduos que são selecionados aleatoriamente. Este vetor-diferença é então somado a um terceiro indivíduo, também selecionado aleatoriamente, de modo a produzir uma solução mutante. A partir deste procedimento, obtém-se então uma população mutante V_t . Os indivíduos da população corrente são então recombinaados com os indivíduos da população mutante, para se pro-

duzir a população de soluções teste U_t . Na versão clássica do DE, emprega-se recombinação discreta, com probabilidade $C \in [0, 1]$. Por fim, compara-se um a um, os indivíduos da população de solução teste com os indivíduos da população corrente, em que sobrevive apenas o mais adaptado.

B. Abordagem para o problema

Para o cálculo da função de custo, em cada avaliação o conjunto de dados é amostrado com 70% para treino e 30% para teste, de forma aleatória. Além disso, em toda geração reavalia-se toda a população, a fim de se evitar a propagação de superindivíduos valorizados pelo fator estocástico da função.

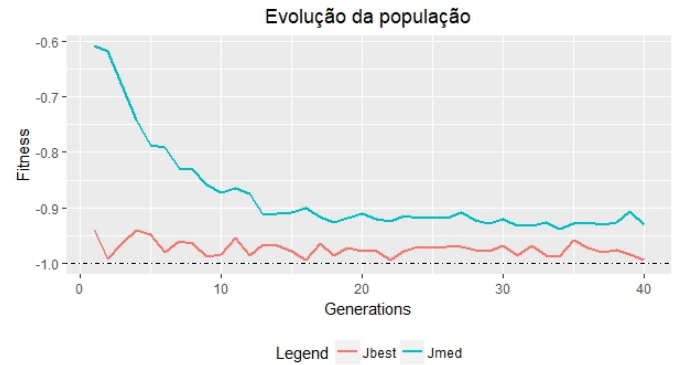


Fig. 2. Evolução ao longo das gerações, melhor e indivíduo médio.

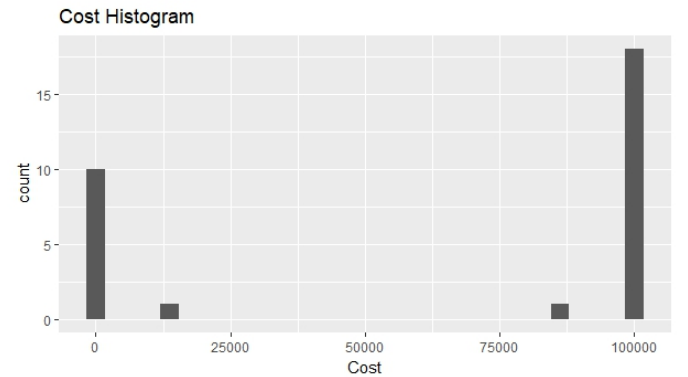


Fig. 3. Distribuição final do atributo custo da população.

É possível observar na Figura 1 que após poucas gerações, para a base de dados *Ionosphere*, toda a população converge para uma região que proporciona uma AUX próxima de 0.9. Além disso, conforme fica evidente na Figura 5, a população final define a região que confere maior eficiência

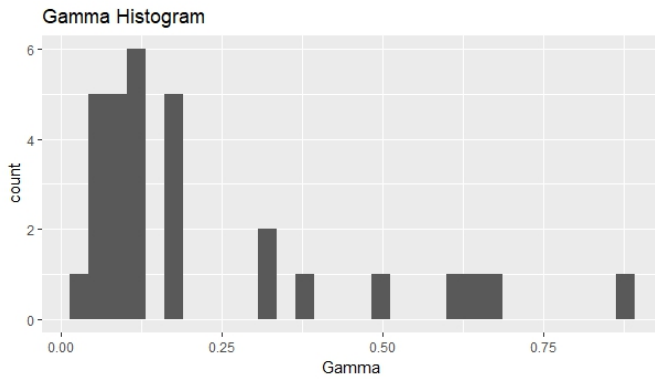


Fig. 4. Distribuição final do atributo gamma da população.

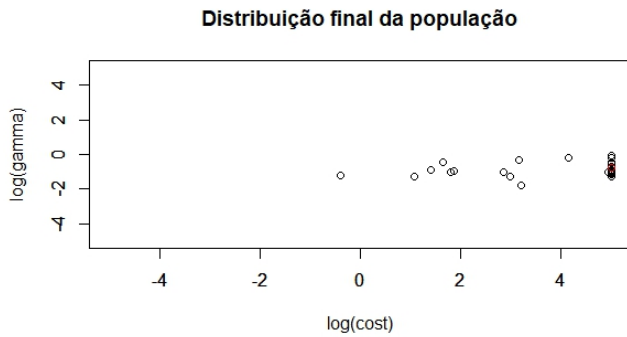


Fig. 5. Distribuição espacial da população final

para a classificação. Em seguida, dado o conjunto de indivíduos, é possível, por exemplo, avaliar toda a população por meio de uma validação cruzada, a fim de determinar o melhor indivíduo da população e a solução ótima para o problema. Na solução proposta, no entanto, foi utilizado o melhor indivíduo ao fim de todas as gerações como solução do problema. Em alguns casos, como na base de dados *BreastCancer* ocorreu de toda a população convergir para o mesmo valor em um dos parâmetros, limitando ainda mais o espaço de busca final e conferindo significância e robustez para o método.

III. RESULTADOS

Para avaliar e comparar os métodos baseados em Evolução Diferencial, i.e DE, e Validação Cruzada, i.e CV, foram utilizadas as bases de dados *Breast Cancer*, *Ionosphere*, *BaseCar*, *Sonar*, *PimaIndiansDiabetes* e *Melanoma*.

Base de dados	DE	CV
Ionosphere	0.9368	0.9253
Breast Cancer	0.9617	0.9685
PimaIndiansDiabetes	0.7110	0.6988
BaseCar	0.9179	0.5000
Sonar	0.8706	0.7427
Melanoma	0.6055	0.5746

TABLE I
AUC DE CLASSIFICAÇÃO PARA AMBOS OS MÉTODOS

IV. CONCLUSÃO

A comparação entre ambos os métodos é realizada apenas em termos de eficiência de classificação, por meio do indicador AUC. Isso porque tempo de execução é um fator relativo à forma como o método é implementado. Sendo assim, conforme fica evidente na Tabela I, o algoritmo diferencial evolutivo, na maioria dos casos, alcança uma eficiência de classificação consideravelmente maior. O fator estocástico da função objetivo provoca uma alteração na convergência montônica do DE, já que a reavaliação do indivíduo provoca uma alteração em seu valor de aptidão. Sendo assim, a população converge para uma região, porém mantém a exploração de modo a retornar parâmetros com maior resolução para γ e custo. Já a validação cruzada, por outro lado, avalia a parametrização do modelo com resolução limitada aos pontos de teste passados como entrada da função e, como o objetivo é explorar uma região consideravelmente larga, estes parâmetros muitas vezes são passados em escala logarítmica. Sendo assim, de acordo com a própria abordagem de otimização, pode-se dizer que o DE realiza uma busca de forma mais inteligente dentro da faixa de valores possíveis do que simplesmente testar possíveis combinações de parâmetros. É interessante perceber que a robustez e capacidade de convergência do DE se mantém inclusive para bases de dados de complexidade elevada e bases desbalanceadas, como é o caso da base *BaseCar*, que atingiu AUC de 0.9179.