

UTILIZAÇÃO DE MÁQUINA DE VETOR DE SUPORTE NA CARACTERIZAÇÃO DE PACIENTES PORTADORES DE DOENÇAS CARDIOVASCULARES.

Juliana B. Azzi¹

Discente do Programa de Pós Graduação em Modelagem Matemática e Computacional - UFRRJ
Italo G. do Vale²

Discente do Programa de Pós Graduação em Ciência, Tecnologia e Inovação Agropecuária - UFRRJ
Robson M. da Silva³

Docente do Programa de Pós Graduação em Modelagem Matemática e Computacional - UFRRJ

1 Introdução

As doenças cardiovasculares são a principal causa de morte no mundo. Segundo [1], estima-se que até o ano de 2030, irão morrer 23,6 milhões de pessoas por causa de doenças cardíacas. No Brasil são responsáveis por cerca de 20% de todos os óbitos em indivíduos com mais de 30 anos. Apesar de boa parte das doenças que envolve o coração poderem ser diagnosticadas com antecedências, o número de casos continua a crescer de forma acentuada principalmente em países pobres e em desenvolvimento. As doenças cardiovasculares são diagnosticadas usando uma disposição de análises laboratoriais e de estudos da imagem lactente. Sendo o Eletrocardiograma (ECG), exame que avalia a propagação elétrica no músculo cardíaco, o principal exame utilizado pelos profissionais de saúde para o diagnóstico de doenças cardíacas nos dias atuais. Apesar da sua eficiência essa técnica não permite incorporar informações laboratoriais. Portanto, torna-se necessário o desenvolvimento de métodos eficientes que auxiliem os profissionais de saúde na tomada de decisão sobre as doenças cardiovasculares. Estudo realizado por [2], obteve acurácia de 83,5% na classificação de pacientes portadores de enfermidades cardíacas (doente e não doente), utilizando a técnicas de redes neurais. Nesse contexto, este trabalho objetiva apresentar um modelo estruturado em *Support Vector Machine* (SVM) de sorte a auxiliar o diagnóstico de indivíduos com doença cardíaca.

¹juliana.baroni@gce.inatel.br

²italovale@ufrj.br

³robsonms@ufrj.br.

2 Materiais e Métodos

O modelo SVM consiste na seleção de um hiperplano que minimize o risco estrutural, a partir da resolução de um problema convexo quadrático. Quando associada à função de kernel possibilita a construção de classificadores não-lineares, através do mapeamento dos dados em um espaço de dimensão superior ao original [3]. Estas características garantem ao SVM uma boa capacidade de categorizar corretamente dados não relacionados na amostra de treinamento. O conjunto de dados utilizado consiste em 270 indivíduos, obtidos da base de dados público (“Heart Disease Database”) [4]. Sendo, 150 indivíduos não-doentes, e o restante, 120, de indivíduos doentes. Os atributos utilizados são listados a seguir: idade, gênero, pressão arterial em repouso, colesterol no soro sanguíneo, concentração de açúcar no sangue, máxima taxa de batimento cardíaco atingida, angina induzida por exercício e diagnóstico. A função de núcleo utilizada foi a RBF (Radial Basis Function) e os parâmetros γ (gamma) e C (custo) foram definidos utilizando validação cruzada no conjunto de treinamento. O modelo SVM foi implementado utilizando o software R e o pacote kernlab [5]. Para avaliar a capacidade de generalização do modelo, foi utilizado a matriz de confusão e o conjunto de dados foi dividido ao mero acaso em dois conjuntos independentes. O conjunto teste composto com 90% do total das amostras do conjunto de dados e o teste com os outros 10%.

3 Resultados e Conclusões

Os resultados obtidos na validação da metodologia proposta neste estudo foram considerados bastante satisfatórios, visto que, a acurácia obtida foi de 82,14% no conjunto de teste. No que tange ao desempenho dos valores preditivo positivo (VPP) e valores preditivo negativo (VPN) o modelo obteve respectivamente 78,9% e 88,89%. Para confirmar a eficácia do modelo, novos testes serão realizados com bases de dados maiores e comparação com a utilização de outros algoritmos de classificação (Naive Bayes e k-Nearest neighbor).

Referências

- [1] ORGANIZATION, W.H. Disponível em: www.who.int/en/. Acessado em 09/2016.
- [2] T. B. Rodrigues, J. L. R. Macrini, E. C. Monteiro. Seleção de Variáveis e Classificação de Padrões por Redes Neurais como auxílio ao Diagnóstico de Cardiopatia Isquêmica *Pesquisa Operacional*, volume 28, n.2, 2008. DOI:10.1590/S0101-74382008000200007.
- [3] B. Schölkopf e A. J. Smola. Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond, *MIT Press.*, Cambridge, 2002.
- [4] D. W. Aha. Heart Disease Databases. Disponível em: www.ics.uci.edu/pub/machine-learning-databases/heart-disease.names. Acessado em 08/2016.
- [5] A. Karatzoglou, A. Smola, K. Hornik. Kernel-Based Machine Learning Lab *Journal of Statistical Software*, volume 11, n.9, 2014.