

Um estudo comparativo de modelos de programação linear para problemas de classificação binária aplicados à diagnósticos de covid

Danielle Vasconcelos Gomes

Faculdade de Ciências Aplicadas - UNICAMP
R. Pedro Zaccaria, 1300, Limeira - SP, 13484-350
d245465@dac.unicamp.br

Cristiano Torezzan

Faculdade de Ciências Aplicadas - UNICAMP
R. Pedro Zaccaria, 1300, Limeira - SP, 13484-350
torezzan@unicamp.br

RESUMO

Um problema central na área de aprendizado de máquina (AM) é a classificação de amostras ou indivíduos em grupos predefinidos, de acordo com as características disponíveis. Quando a aplicação envolve dois grupos, esse problema é conhecido como classificação supervisionada binária e tem muitas aplicações, incluindo modelos para previsão de diagnósticos médicos. Várias abordagens têm sido propostas para modelar esse problema, incluindo métodos avançados de aprendizado de máquina, como as redes neurais artificiais, florestas aleatórias e gradiente boosting, além de métodos estatísticos que dependem de hipóteses sobre a distribuição dos dados. De forma alternativa, modelos de programação linear têm permitido formulações de classificação binária motivados pela ideia de encontrar um hiperplano separador dos dados. Essa abordagem é a base de métodos como máquina de vetores de suporte (SVM) e formulações de programação por metas para classificação binária. Mais recentemente, um novo método de Análise Discriminante por Programação Linear (LDPA) tem obtido resultados promissores. Neste trabalho comparamos o desempenho de métodos baseados em programação linear para o problema de classificação binária aplicado ao diagnóstico de covid com base em sintomas relatados pelos pacientes. O estudo de base empírico-computacional utiliza um conjunto de dados reais anonimizados de 64mil pacientes brasileiros. Os resultados mostram que métodos baseados em PL, sobretudo LDPA, podem atingir desempenho compatíveis com algoritmos clássicos de aprendizado de máquina.

PALAVRAS CHAVE. Programação linear, Classificação binária, Diagnóstico médico.

Tópicos: SA – PO na Área de Saúde; PM – Programação Matemática.