No SAS, a função implementada para realizar análises utilizando algoritmos de suporte vetorial (SVM) é o PROC HPSVM. Essa função permite o uso de kernels lineares e não lineares nos dados de treinamento. O PROC HPSVM executa o algoritmo SVM de alta performance, possibilitando sua execução em computação paralela tanto em uma única máquina quanto em múltiplas máquinas.

Utilizando o exemplo fornecido na documentação do SAS com um kernel linear, vamos ajustar um modelo utilizando um banco de dados chamado SAMPSIO.DMAGECR, um banco de dados de referência que traz informações sobre risco de crédito. Esse banco de dados faz parte da biblioteca SAMPSIO e contém 1.000 observações, cada uma com informações detalhadas sobre os requerentes. O banco inclui a classificação do indivíduo como GOOD ou BAD em uma variável denominada GOOD\_BAD, além de outras variáveis como histórico de crédito, duração do empréstimo, entre outras.

*Tabela

Descrição gerada automaticamente*

A tabela 'Matriz de Classificação' mostra que, entre as 1.000 observações totais, 700 são classificadas como boas e 300 como ruins. O número de observações BOAS corretamente previstas é 626, e o número de observações RUINS corretamente previstas é 158.

Tabela

Descrição gerada automaticamente

Assim, a precisão do modelo é de 78,4%, conforme indicado na tabela 'Estatísticas de Ajuste'.

Tabela

Descrição gerada automaticamente

Um modelo relativamente bom significa que a taxa de erro de classificação é baixa, enquanto a sensibilidade e a especificidade são altas. Com o PROC HPSVM, é possível ajustar os parâmetros de treinamento e usar diferentes kernels para obter um modelo melhor. O padrão do procedimento usa o kernel linear conforme especificado:

\begin{aling} k(x\_1, x\_2) = \langle x\_1, x\_2 \rangle \end{ aling }

onde $x\_1$ e $x\_2$ são dois vetores e $ \langle \cdot, \cdot \rangle $é o produto interno.

Para alterar o tipo de kernel utilizado, basta definir o argumento KERNEL, especificando o tipo de kernel e quaisquer parâmetros associados. Definindo o kernel como polinomial:

\begin{align}

k(x\_1, x\_2) = (\langle x\_1, x\_2 \rangle + 1)^p

\end{align}

onde $p$ é o grau do polinômio, obtém-se o resultado:

Tabela

Descrição gerada automaticamente

A matriz de classificação mostra que o número de observações BOAS corretamente previstas é 699, e o número de observações RUINS corretamente previstas é 297. Assim, a precisão do modelo é de 99,6%, conforme indicado na tabela 'Estatísticas de Ajuste', mostrando que o uso do kernel polinomial é mais acurado, nesse caso.

Tabela

Descrição gerada automaticamente