



Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



REDES NEURAIS ARTIFICIAIS REGULARIZADAS PARA PREDIÇÃO GENÔMICA DA RESISTÊNCIA A CERCOSPORIOSE EM *COFFEA ARABICA*

Noé Mitterhofer Eiterer Ponce de Leon da Costa¹, Moysés Nascimento¹, Victória Manhago Salvador¹, Eveline Teixeira Caixeta¹, Antonio Carlos Baião de Oliveira¹

¹Universidade Federal de Viçosa, Centro de Ciências Exatas, Departamento de Estatística, Viçosa, Minas Gerais.

Autor correspondente: noe.costa@ufv.br

Palavras-Chaves: seleção genômica, inteligência artificial, overfitting

Área temática: Inteligência Artificial; Grande Área: Ciências Agrárias; Categoria do Trabalho: Pesquisa

Introdução

Pouco explorado pelas iniciativas privadas, o melhoramento do café é mais desenvolvido pelas instituições públicas. Uma das formas de se acelerar o processo de identificar variedades com características agronomicamente superiores é através da Seleção Genômica (SG). A SG basicamente utiliza das informações moleculares para prever o desempenho dos genótipos avaliados. Entre vários métodos utilizados para predição genômica, as Redes Neurais Artificiais (RNA) apresentam a grande vantagem de conseguir capturar, a partir dos marcadores moleculares, as relações lineares e não-lineares de maneira natural sem a necessidade de se estabelecer um modelo a priori. Porém, as RNA podem sofrer com o *overfitting*, ou seja, a rede memoriza aquele conjunto de dados e perde sua capacidade preditiva em outros conjuntos de dados. Uma forma de evitar o *overfitting* é através do uso de RNA Regularizadas (RNAR), como por exemplo a do tipo L1. A regularização L1 basicamente modifica a função de custo da rede para penalizar grandes pesos e tendendo a fazer com que a rede prefira pequenos pesos.

Objetivos

O objetivo desse trabalho foi de comparar a capacidade preditiva de diferentes arquiteturas de RNAR do tipo L1 na predição do mérito de genótipos quanto a resistência a cercosporiose em *Coffea arabica*.

Material e Métodos

Os dados utilizados foram de resistência a cercosporiose de 156 genótipos de *Coffea arabica* genotipados para 3824 marcadores. Do conjunto total de dados (156 indivíduos), 80% foram utilizados para treinamento da rede e o restante para teste.

Foram utilizadas RNA do tipo Perceptron de Múltiplas Camadas, com duas camadas ocultas, cada uma variando 1 a 20 neurônios. Foram consideradas 50 iterações e valores do parâmetro de regularização (k) iguais à 2, 4, 8 ou 10. Como métrica de comparação foi utilizada a capacidade preditiva da rede.

Resultados e Discussão

Considerando $k=2$ a maior capacidade preditiva foi de 0,12 com 8 neurônios na primeira camada e 1 neurônio na segunda camada. Com $k=4$, o resultado obtido de capacidade preditiva foi de 0,35 quando utilizou 3 neurônios na primeira e 8 na segunda camada. Já considerando $k=8$ regularizações, o maior valor de capacidade preditiva da rede foi de 0,51 quando se utilizou 11 neurônios na primeira camada e 15 na segunda camada oculta. Finalmente, com $k=10$, a maior capacidade preditiva foi de 0,44 com 2 neurônios em ambas as camadas.

Conclusões

Conclui-se que mantendo o número de iterações em 50, podemos verificar que o aumento de regularizações em até 8 resultou em aumento da capacidade preditiva da rede. Isso pode ser explicado pela capacidade das regularizações em evitar o *overfitting*. Elevados valores do parâmetro de regularização, neste trabalho, igual à 8 ocorre redução da capacidade preditiva.

Apoio Financeiro



Agradecimentos

