



XIII SIGM

International symposium on
genetics and breeding

UMA ABORDAGEM ENSEMBLE PARA SPLINES DE REGRESSÃO ADAPTATIVA MULTIVARIADA EM ESTUDOS DE PREDIÇÃO GENÔMICA

XIII International Symposium on Genetics and Breeding, 13ª edição, de 25/10/2022 a 27/10/2022
ISBN dos Anais: 978-65-5465-014-4

CELERI; Maurício de Oliveira¹, COSTA; Weverton Gomes da², NASCIMENTO; Ana Carolina Campana³, AZEVEDO; Camila Ferreira⁴, NASCIMENTO; Moysés⁵

RESUMO

Splines de Regressão Adaptativa Multivariada (MARS) é uma metodologia de aprendizado estatístico que vem ganhando espaço representativo dentro de estudos de predição e seleção genômica. A MARS apresenta uma característica interessante de selecionar poucas variáveis de entrada (em geral, marcadores moleculares do tipo SNP) para compor o modelo final de predição, tornando-a notavelmente eficiente para características controladas por poucos genes de grandes efeitos, no entanto, seu desempenho para características poligênicas, na maioria das vezes, se torna limitado. O objetivo deste trabalho é avaliar uma proposta de um método *ensemble* homogêneo para a MARS inspirada na proposta do *random forest*. A hipótese é de que esta proposta pode tornar a MARS uma metodologia com resultados tão satisfatórios para características poligênicas quanto para características oligogênicas, aumentando sua capacidade preditiva e o número de marcadores selecionados. O método proposto consiste em amostrar B vezes, aleatoriamente, um subconjunto de tamanho M dos P marcadores disponíveis; em cada uma das B amostragens um modelo MARS é estimado gerando um total de B predições distintas para a característica de interesse, de forma que a predição final da metodologia proposta é a média aritmética simples entre estas B predições. Para a avaliação desta proposta utilizou-se 12 cenários simulados, definidos pela combinação de quatro níveis de números de QTLs (8, 40, 80 e 120) e três níveis de herdabilidade (0,3, 0,5 e 0,8) com 4010 marcadores distribuídos em 10 grupos de ligação. Para todos os cenários simulados foram considerados efeitos aditivos e epistáticos e gmd igual a 0,5. Para a validação utilizou-se da validação cruzada 10-folds e os resultados de predição do método proposto foi comparado com o modelo MARS e o modelo GBLUP considerando efeitos aditivos e devidos a dominância e à epistasia. Para a avaliação do método proposto considerou-se a construção de 50 modelos MARS com a amostragem de 500 dos 4010 marcadores por vez. Em todos os cenários o método proposto obteve capacidade preditiva superiores à MARS, em relação ao GBLUP observou-se resultados de capacidade preditiva superior para o método proposto nos cenários controlados por 8 QTLs ou em herdabilidade 0,8, para os demais cenários obteve-se resultados similares. O EQM do método proposto sempre esteve abaixo ou

¹ Universidade Federal de Viçosa, mauricio.celeri@ufv.br

² Universidade Federal de Viçosa, wevertonufv@gmail.com

³ Universidade Federal de Viçosa, ana.campana@ufv.br

⁴ Universidade Federal de Viçosa, camila.azevedo@ufv.br

⁵ Universidade Federal de Viçosa, moysesnascim@ufv.br

comparável às demais metodologias. Em contrapartida, o tempo computacional foi superior aos demais dois métodos, gastando em média 96s para o ajuste enquanto MARS e GBLUP gastaram em média 31s e 18s, respectivamente. O método proposto seleciona, em média, 480 dos 4010 marcadores, número este que chega a ser até 30 vezes maior do que o número de marcadores selecionados pelo modelo MARS. A proposta do método *ensemble* para a MARS obteve resultados satisfatórios nos cenários avaliados, mostrando que a MARS tem potencial para trabalhar com características quantitativas tão bem quanto outras abordagens.

PALAVRAS-CHAVE: método ensemble, MARS, características quantitativas