

Confirmation

You abstract has been submitted

Your Abstract ID is: **309704**

Submission Type: Refereed

If abstract cannot be accomodated for a paper presentation, would like to be considered for a poster presentation: No

Topic: Machine Learning

Title: Quantifying the bias-variance decomposition using metaled distributions

Abstract Text:

The bias-variance decomposition (BVD) serves as a key guidepost in machine-learning practice. Comparing model scores for training, validation, and test data provide qualitative indications of potential model overfitting. Explicit calculation of BVD terms is generally problematic, absent assumptions about convenient probability distributions. Metalog distributions, an emerging technique from the Decision-Analysis community, provide a mechanism for explicit calculation of BVD terms. Metalogs fit arbitrary empirical distributions to closed-form, multiply-differentiable, continuous functions represented as series. The bias and variance terms result from metalog-distribution fits to prediction-model training- and test-data residual errors employing a method resembling Locally Weighted Scatterplot Smoothing (LOWESS). Instead of calculating means as is done with LOWESS, metalog-fit probability-density functions are instead obtained. This allows for some statistical inference using machine-learning results, including estimates of prediction and confidence intervals. Using Kolmogorov-Smirnov or other tests to the resulting bias and variance distributions leads to a quantitative characterization of prediction-model overfitting. Metalog distributions can similarly be employed to estimate Cramer-Row Lower Bounds (CRLB) on estimation-error variance. This occurs from fitting orthogonal-transform-space representations of the explanatory variables to metaled distributions. The CLRB offers a proxy for the irreducible-variance term in the BVD. Explicit calculation of BVD terms using metalog distributions affords the opportunity to extend Analysis of Variance (ANOVA) methods from classical statistics into the machine-learning realm.

Keywords: machine learning, mathematical statistics, probability distributions

1: Hamlett, Neil - Uncertainty Research, LLC (presenting)

[Enter another abstract](#)

[2021 Symposium on Data Science & Statistics Homepage](#)