

Thesis proposal: Smooth terms in generalized additive models

Raul Kangro

Tartu, Fall 2025

Introduction

In machine learning practice the approach allowing very flexible fit to prediction function and avoiding overfitting by some kind of regularization have proved to be very successful. Neural networks are most widely used methods corresponding to this approach but often alternatives like generalized additive models work much faster and provide comparable accuracy if used efficiently. Efficiency means that correct type of smoothing terms with suitable penalty functions should be used.

Unfortunately, in general machine learning courses there is no time to discuss the meaning of various smoothing terms in detail, therefore there is a need for supplementary study materials with appropriate examples to support independent study in this area.

Description of a proposed thesis work

The thesis should provide definitions and well chosen illustrations of various versions of spline functions used in `mgcv` package of R programming language together with effects of different penalty terms on the result of fit. The student should provide clear mathematical descriptions of the functions and work out original examples demonstrating potential usefulness of various terms in certain situations. The last chapter of the thesis should contain an example of applying GAM methods (with discussion how suitable smoothing terms were selected) to a practical data set and comparing the results to the ones obtained by some alternative methods like random forests or linear models without smoothing terms.