ABSTRACT

Generalized Additive Models (GAM), introduced by Hastie and Tibshirani (1990), are a nice generalization of logistic and linear regression. We give the first efficient estimators for these models based on a modification of Regression Trees.

More precisely, we assume that the data is distributed according to a GAM:

$$E[y|x] = u(f_1(x_1) + f_2(x_2) + ... + f_n(x_n)),$$

where the function $u$ is a monotonic function and each $f_i$ is an arbitrary function of the $i^{th}$ attribute of $x$. The estimator is efficient in two senses: it outputs a function $h(x)$ with mean squared error $< \epsilon$ for arbitrarily small $\epsilon > 0$, and the amount of data and computational runtime required of the learning algorithm are polynomial in $1/\epsilon$, the dimension of the problem $n$, and bounds on the derivatives of the functions.