

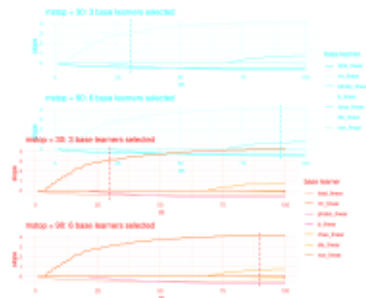
Introduction to Machine Learning



Boosting Boosting: Advanced CWB

Gradient Boosting: Advanced CWB

Learning goals



- Details of nonlinear BLs and splines
- Decomposition for splines
- Fair base learner selection
- Details of nonlinear BLs and splines
- Feature importance and PDPs
- Decomposition for splines
- Fair base learner selection
- Feature importance and PDPs

NONLINEAR EFFECT DECOMPOSITION

► Kneib, Hothorn, and Tibshirani (2009) proposed a decomposition of each base learner into a constant, a linear and a nonlinear part. The boosting algorithm will automatically decide which feature to include: linear, nonlinear, or none at all:

$$\begin{aligned} b_j(x_j, \theta^{[m]}) &= b_{j,\text{const}}(x_j, \theta^{[m]}) + b_{j,\text{lin}}(x_j, \theta^{[m]}) + b_{j,\text{nonlin}}(x_j, \theta^{[m]}) \\ &= \theta_{j,\text{const}}^{[m]} + x_j \cdot \theta_{j,\text{lin}}^{[m]} + s_j(x_j, \theta_{j,\text{nonlin}}^{[m]}), \end{aligned}$$

where

- $\theta_{j,\text{const}}$ is the intercept of feature j ,
- $x_j \cdot \theta_{j,\text{lin}}^{[m]}$ is a feature-specific linear base learner, and
- $s_j(x_j, \theta_{j,\text{nonlin}}^{[m]})$ is a (centered) nonlinear base learner capturing deviation from the linear effect

Careful: We usually also apply an orthogonalization procedure on top of this but skip technical details here.

