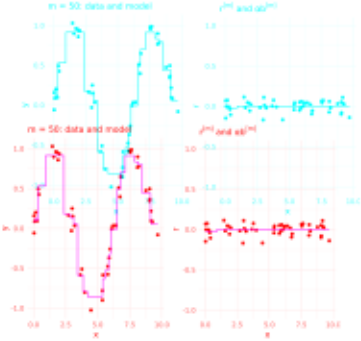
Introduction to Machine Learning

Boosting

Gradient Boosting with Trees 1

Gradient Boosting with Trees 1



Learning goals

- Examples for GB with trees
- Understand relationship between model structure and interaction depth

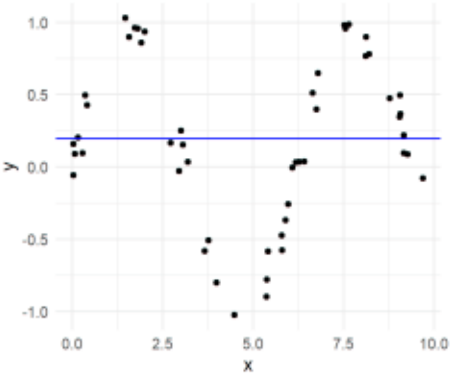
Learning goals

- Examples for GB with trees
- Understand relationship between model structure and interaction depth

EXAMPLE 1 / 2

Iteration 0: initialization by optimal constant (mean) prediction $\hat{f}^{[0]}(x^{(i)}) = \bar{y} \approx 0.2$.

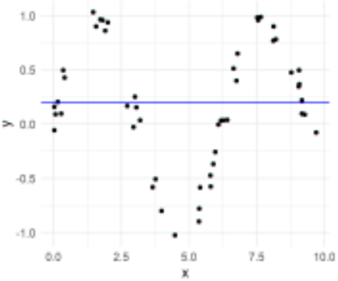
i	$x^{(i)}$	$y^{(i)}$	$\hat{f}^{[0]}$
1	0.03	0.16	0.20
2	0.03	-0.06	0.20
3	0.07	0.09	0.20
$\vdots$	$\vdots$	$\vdots$	$\vdots$
50	9.69	-0.08	0.20



EXAMPLE 1 / 3

Iteration 1: (1) Calculate pseudo-residuals $\tilde{r}^{[m](i)}$ and (2) fit a regression stump $b^{[m]}$.

i	$x^{(i)}$	$y^{(i)}$	$\hat{f}^{[0]}$	$\tilde{r}^{[1](i)}$	$\hat{b}^{[1](i)}$
1	0.03	0.16	0.20	-0.04	-0.17
2	0.03	-0.06	0.20	-0.25	-0.17
3	0.07	0.09	0.20	-0.11	-0.17
⋮	⋮	⋮	⋮	⋮	⋮
50	9.69	-0.08	0.20	-0.27	0.33



(3) Update model by $\hat{f}^{[1]}(x) = \hat{f}^{[0]}(x) + \hat{b}^{[1]}$.

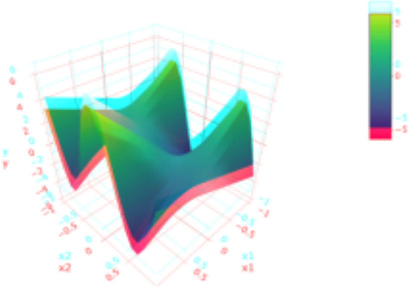
MODEL STRUCTURE AND INTERACTION DEPTH / 2

Simulation setting:

- Features x_1 and x_2 and numeric y; with $n = 500$
- x_1 and x_2 are uniformly distributed between -1 and 1
- $y^{(i)} = x_1^{(i)} - x_2^{(i)} + 5\cos(5x_2^{(i)}) \cdot x_1^{(i)} + \epsilon^{(i)}$ with $\epsilon^{(i)} \sim \mathcal{N}(0, 1)$
- We fit 2 GB models, with tree depth 1 and 2, respectively.

Simulation setting:

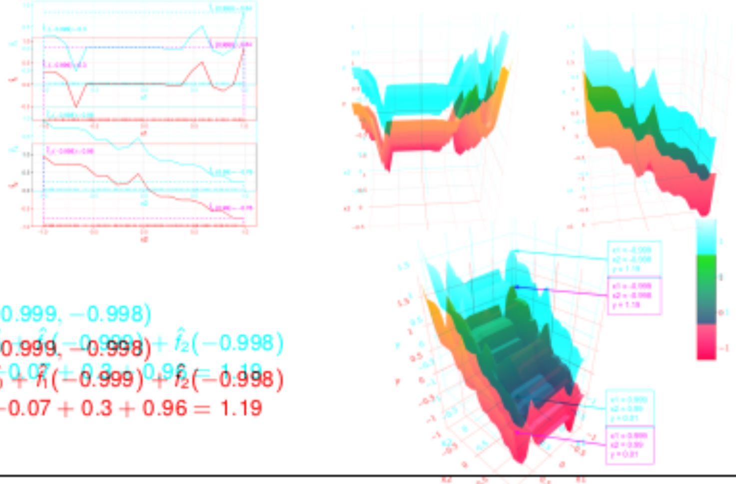
- Features x_1 and x_2 and numeric y; with $n = 500$
- x_1 and x_2 are uniformly distributed between -1 and 1
- $y^{(i)} = x_1^{(i)} - x_2^{(i)} + 5\cos(5x_2^{(i)}) \cdot x_1^{(i)} + \epsilon^{(i)}$ with $\epsilon^{(i)} \sim \mathcal{N}(0, 1)$
- We fit 2 GB models, with tree depth 1 and 2, respectively.



MODEL STRUCTURE AND INTERACTION DEPTH / 3

GBM with interaction depth of 1 (GAM)

No interactions are modelled: Marginal effects of x_1 and x_2 add up to joint effect (plus the constant intercept $\hat{f}_0 = -0.07$).



$$\hat{f}(-0.999, -0.998) = \hat{f}_0 + \hat{f}_1(-0.999) + \hat{f}_2(-0.998) = -0.07 + 0.3 + 0.96 = 1.19$$

MODEL STRUCTURE AND INTERACTION DEPTH / 4

GBM with interaction depth of 2

Interactions between x_1 and x_2 are modelled: Marginal effects of x_1 and x_2 do NOT add up to joint effect due to interaction effects.

GBM with interaction depth of 2

Interactions between x_1 and x_2 are modelled: Marginal effects of x_1 and x_2 do NOT add up to joint effect due to interaction effects.