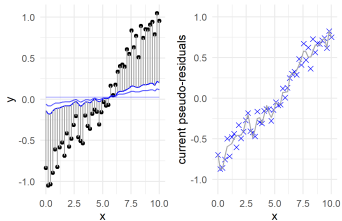


Gradient Boosting: Illustration

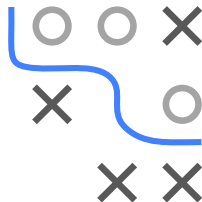


- See simple visualizations of boosting in regression
- Understand impact of different losses and base learners

GRADIENT BOOSTING ILLUSTRATION - GAM

GAM / Splines as BL and compare L_2 vs. L_1 loss.

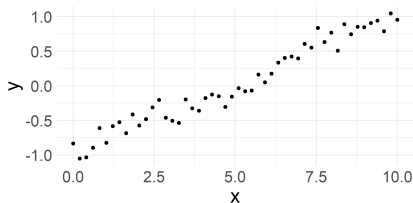
- L_2 : Init = optimal constant = $\text{mean}(y)$; for L_1 it's $\text{median}(y)$
- BLs are cubic B -splines with 40 knots.
- PRs L_2 : $\tilde{r}(f) = r(f) = y - f(\mathbf{x})$
- PRs L_1 : $\tilde{r}(f) = \text{sign}(y - f(\mathbf{x}))$
- Constant learning rate 0.2



Univariate toy data:

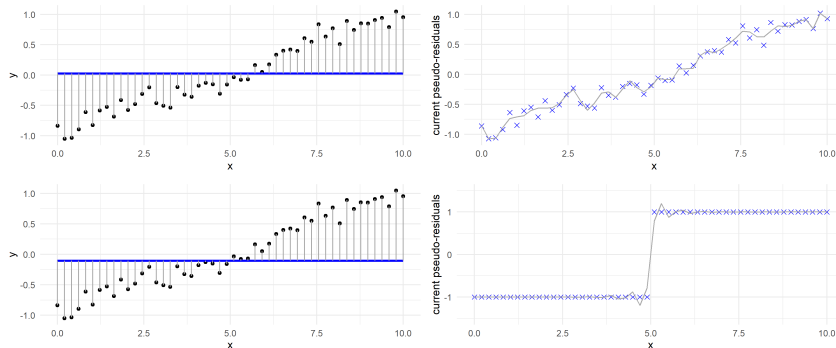
$$y^{(i)} = -1 + 0.2 \cdot x^{(i)} + 0.1 \cdot \sin(x^{(i)}) + \epsilon^{(i)}$$

$$n = 50 ; \epsilon^{(i)} \sim \mathcal{N}(0, 0.1)$$



GAM WITH L_2 VS L_1 LOSS

Top: L_2 loss, bottom: L_1 loss

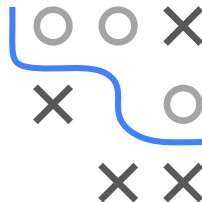
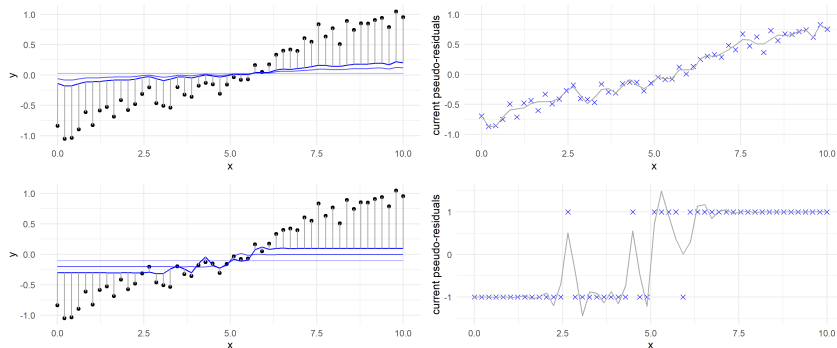


Iteration 1

Shape of PRs affects gradual model fit: L_1 only sees residuals' sign, BLs are not affected size of values as in L_2 and hence lead to more moderate changes.

GAM WITH L_2 VS L_1 LOSS

Top: L_2 loss, bottom: L_1 loss

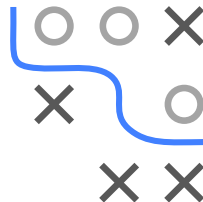
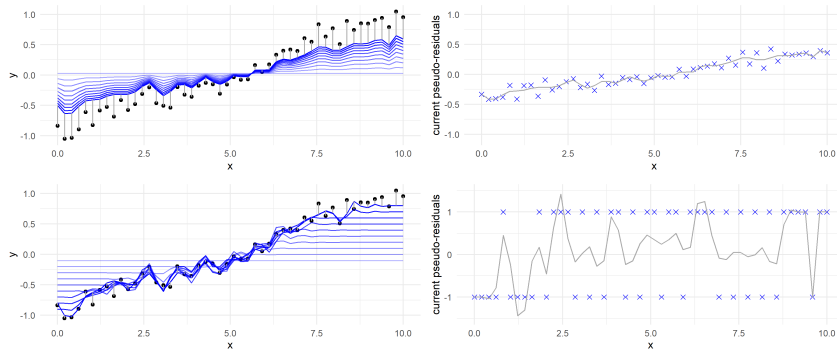


Iteration 3

Shape of PRs affects gradual model fit: L_1 only sees residuals' sign, BLs are not affected size of values as in L_2 and hence lead to more moderate changes.

GAM WITH L_2 VS L_1 LOSS

Top: L_2 loss, bottom: L_1 loss

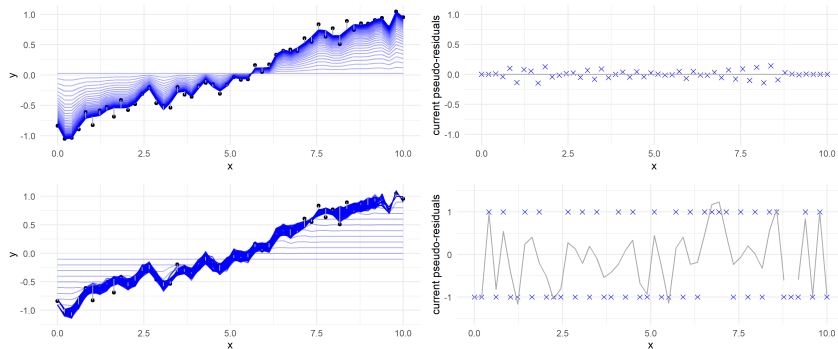


Iteration 10

Shape of PRs affects gradual model fit: L_1 only sees residuals' sign, BLs are not affected size of values as in L_2 and hence lead to more moderate changes.

GAM WITH L_2 VS L_1 LOSS

Top: L_2 loss, bottom: L_1 loss

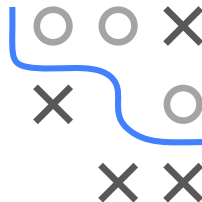
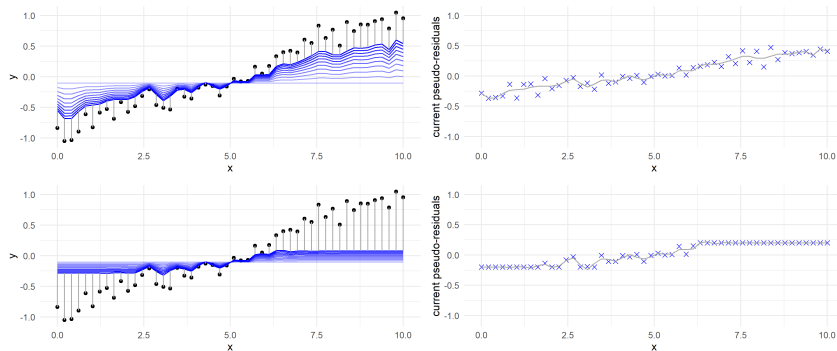


Iteration 100

Shape of PRs affects gradual model fit: L_1 only sees residuals' sign, BLs are not affected size of values as in L_2 and hence lead to more moderate changes.

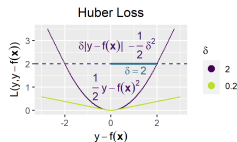
GAM WITH HUBER LOSS

Top: $\delta = 2$, bottom: $\delta = 0.2$.



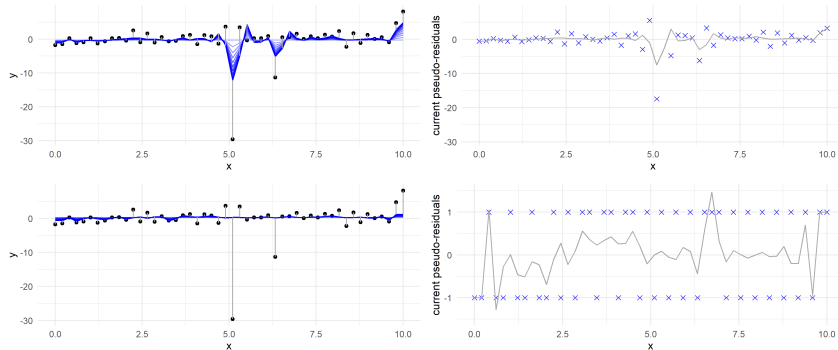
Iteration 10

For small δ , PRs are often bounded, resulting in $L1$ -like behavior, while the upper plot more closely resembles $L2$ loss.



GAM WITH OUTLIERS

Instead of Gaussian noise, let's use t -distrib, that leads to outliers in y .
Top: $L2$, bottom: $L1$.



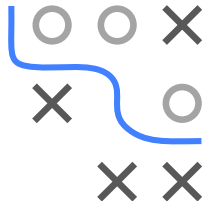
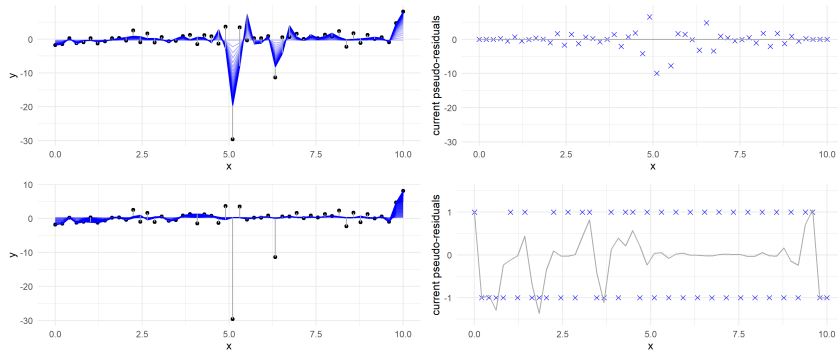
Iteration 10

$L2$ loss is affected by outliers rather strongly, whereas $L1$ solely considers residuals' sign and not their magnitude, resulting in a more robust model.



GAM WITH OUTLIERS

Instead of Gaussian noise, let's use t -distrib, that leads to outliers in y .
Top: $L2$, bottom: $L1$.

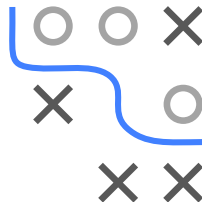
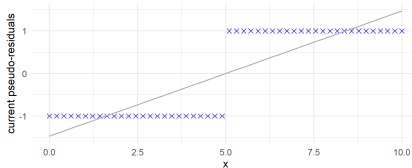
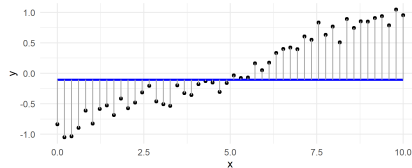
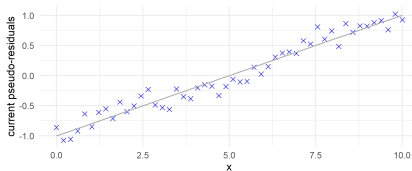
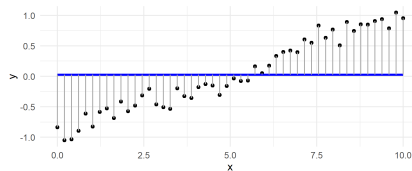


Iteration 100

$L2$ loss is affected by outliers rather strongly, whereas $L1$ solely considers residuals' sign and not their magnitude, resulting in a more robust model.

LM WITH L_2 VS L_1 LOSS

Top: L_2 , bottom: L_1 .

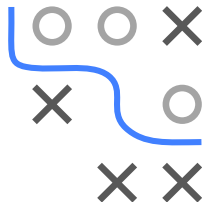


Iteration 1

L_2 : as $\tilde{r}(f) = r(f)$, BL of 1st iter already optimal; but learn rate slows us down.

Top: $L2$, bottom: $L1$.

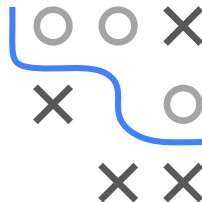
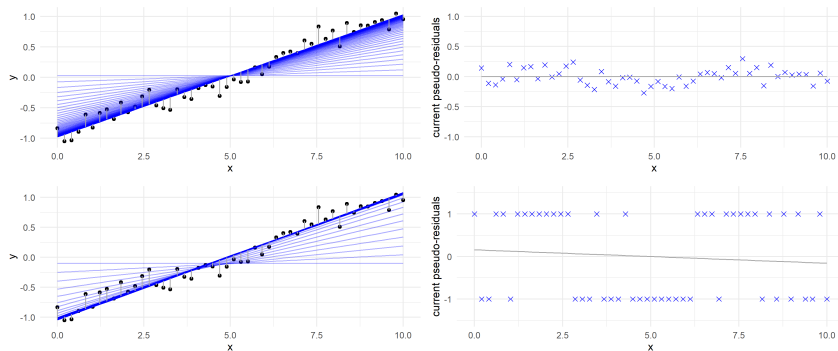
Top: $L2$, bottom: $L1$.



L2: as $\tilde{r}(f) = r(f)$, BL of 1st iter already optimal; but learn rate slows us down.

LM WITH L_2 VS L_1 LOSS

Top: L_2 , bottom: L_1 .



Iteration 100

L_2 : as $\tilde{r}(f) = r(f)$, BL of 1st iter already optimal; but learn rate slows us down.