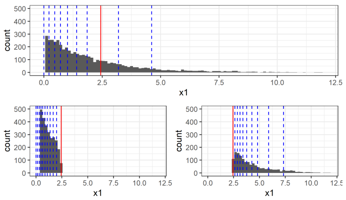
Introduction to Machine Learning

Boosting
Gradient Boosting: Deep Dive XGBoost
Optimization



Learning goals

- Understand details of the regularized risk in XGBoost
- Understand approximation of loss used in optimization
- Understand split finding algorithm

RISK MINIMIZATION

XGBoost uses a risk function with 3 regularization terms:

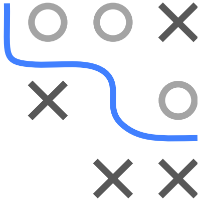
$$
\begin{aligned}
\mathcal{R}_{\text{reg}}^{[m]} = \sum_{i=1}^{n} L &\left(y^{(i)}, f^{[m-1]}(\mathbf{x}^{(i)}) + b^{[m]}(\mathbf{x}^{(i)})\right) \\
&+ \lambda_1 J_1(b^{[m]}) + \lambda_2 J_2(b^{[m]}) + \lambda_3 J_3(b^{[m]}),
\end{aligned}
$$

with $J_1(b^{[m]}) = T^{[m]}$ the number of leaves in the tree to penalize tree depth.

$J_2(b^{[m]}) = \left\|\mathbf{c}^{[m]}\right\|_2^2$ and $J_3(b^{[m]}) = \left\|\mathbf{c}^{[m]}\right\|_1$ are $L2$ and $L1$ penalties of the terminal region values $c_t^{[m]}, t = 1, \dots, T^{[m]}$.

We define $J(b^{[m]}) := \lambda_1 J_1(b^{[m]}) + \lambda_2 J_2(b^{[m]}) + \lambda_3 J_3(b^{[m]})$.

LOSS MINIMIZATION - SPLIT FINDING



To evaluate the performance of a candidate split that divides the instances in region $R_t^{[m]}$ into a left and right node we use the **risk reduction** achieved by that split:

$$\tilde{S}_{LR} = \frac{1}{2}\left[\frac{t_{\lambda_3}\left(G_{tL}^{[m]}\right)^2}{H_{tL}^{[m]} + \lambda_2} + \frac{t_{\lambda_3}\left(G_{tR}^{[m]}\right)^2}{H_{tR}^{[m]} + \lambda_2} - \frac{t_{\lambda_3}\left(G_{t}^{[m]}\right)^2}{H_{t}^{[m]} + \lambda_2}\right] - \lambda_1,$$

where the subscripts L and R denote the left and right leaves after the split.