

Gradient Boosting Trees: The Math



$$0.01 \text{ tree} + 0.01 \text{ tree} + 0.01 \text{ tree} + \dots + 0.01 \text{ tree}$$
$$\textcircled{z_0} + 0.01 \textcircled{z_1} + 0.01 \textcircled{z_2} + \dots + 0.01 \textcircled{z_n}$$

baseline prediction *how do we choose these errors to correct?*

future trees predict error for a regression tree given defined loss function

let F_i be our predictions $F_i = \sum_{t=0}^i z_t$

$$\begin{aligned} F_1 &= z_0 + z_1 \\ F_2 &= z_0 + z_1 + z_3 \\ &\vdots \end{aligned}$$

$$F_i = F_{i-1} + \textcircled{z_i} \qquad z_i = - \frac{\partial \text{Loss}(y, F_i)}{\partial F_i}$$

Gradient Boosting Trees: The Math



$$z_i = - \frac{\partial \text{Loss}(y, F_i)}{\partial F_i}$$

Negative Gradient of Loss w.r.t. Ensemble Prediction

- The Negative Gradient tell us what adjustments we should make to our prediction F_i in order to decrease our loss
- Example:

$$\text{Loss}(y, \hat{y}) = (y - \hat{y})^2 \quad \Longrightarrow \quad -\frac{\partial \text{Loss}(y, \hat{y})}{\partial \hat{y}} \quad \Longrightarrow \quad 2(y - \hat{y})$$

- With squared loss, error is the negative gradient, but the negative gradient will work in other situations!