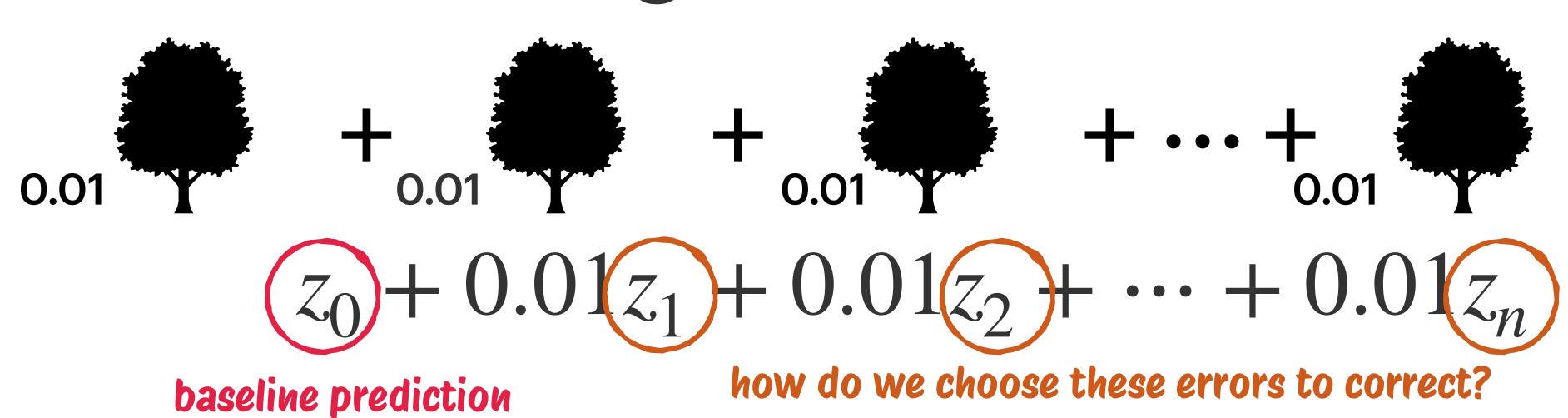
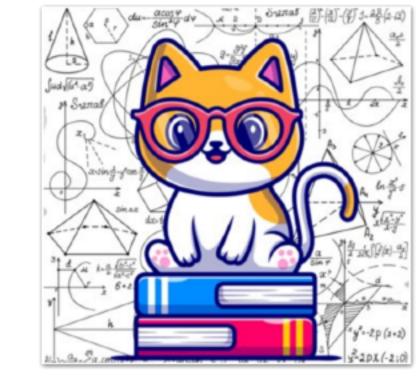
Gradient Boosting Trees: The Math





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

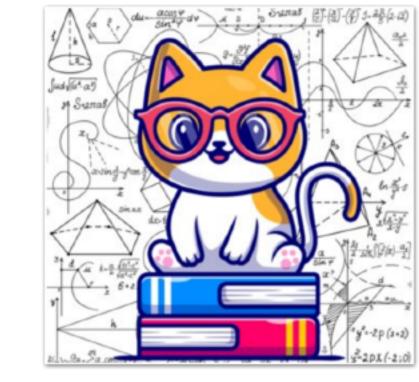
$$\text{let } F_i \text{ be our predictions } F_i = \sum_{t=0}^l z_t \qquad \begin{matrix} F_1 = z_0 + z_1 \\ F_2 = z_0 + z_1 + z_3 \\ \vdots \end{matrix}$$

$$F_i = F_{i-1} + (z_i)$$

$$z_{i} = -\frac{\partial Loss(y, F_{i})}{\partial F_{i}}$$

Gradient Boosting Trees: The Math

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



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

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

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

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