Loss Math

For a given leaf and split, we have that minimum loss is given by

$$\min_{\gamma} \sum_{i=1}^{n} \mathcal{L}(y_i, \frac{\alpha o_i + \gamma}{\alpha + 1})$$

We can set the "plain optimum" to be

$$\gamma^* = \arg\min_{\gamma} \sum_{i \in \text{leaf}} \mathcal{L}(y_i, \gamma)$$

And the "ensemble optimum" to be

$$\tilde{\gamma}^* = \arg\min_{\gamma} \sum_{i \in \text{leaf}} \mathcal{L}(y_i, \frac{\alpha o_i + \gamma}{\alpha + 1})$$

Letting a star denote the optimal value that combines ensemble predictions, I split at A^* and define the the left leaf value as

$$B_{\text{leaf}} = (1 - \delta)(\gamma^*) + \delta \tilde{\gamma}^*$$

With a split at A^*

The split is defined at:

$$\arg\min_{A} \sum_{i=1}^{n} \mathcal{L}(y_i, \frac{\alpha o_i + (1-\delta)\sum_{i=1}^{n} \arg\min_{\gamma} \mathcal{L}(y_i, \gamma) + \delta \sum_{i=1}^{n} \arg\min_{\gamma} \mathcal{L}(y_i, \frac{\alpha o_i + \gamma}{\alpha + 1})}{\alpha + 1})$$

Where we define

$$\sum_{i=1}^{n} (1-\delta) \arg\min_{\gamma} \mathcal{L}(y_i, \gamma) + \sum_{i=1}^{n} \delta \arg\min_{\gamma} \mathcal{L}(y_i, \frac{\alpha o_i + \gamma}{\alpha + 1}) = B_n$$

Generic Algorithm

Begin with initial value

$$\gamma_0 = \arg\min_{\gamma} \sum_{i=1}^{n} \mathcal{L}(y_i, \gamma)$$

Find error at split A

$$\gamma^* = \gamma_0 - \eta \frac{G}{H + \lambda}$$

Given

$$\mathcal{L}(y_i, \frac{\alpha o_i + \gamma_0}{\alpha + 1})$$

We have that $g_i =$

$$\frac{1}{\alpha + 1} \frac{\partial \mathcal{L}(y_i, z)}{\partial z}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}}$$

And $h_i =$

$$\frac{1}{(\alpha + 1)^2} \frac{\partial^2 \mathcal{L}(y_i, z)}{\partial z^2}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}}$$

So that $G = \sum_{i \in \text{leaf}} g_i$ and $H = \sum_{i \in \text{leaf}} h_i$

So, we have that

$$\begin{aligned} \frac{G}{H} &= \\ \frac{\sum_{i \in \text{leaf}} g_i}{\sum_{i \in \text{leaf}} h_i} &= \\ \frac{\sum_{i \in \text{leaf}} \frac{1}{\alpha + 1} \frac{\partial \mathcal{L}(y_i, z)}{\partial z}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}}}{\sum_{i \in \text{leaf}} \frac{1}{(\alpha + 1)^2} \frac{\partial^2 \mathcal{L}(y_i, z)}{\partial z^2}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}}} &= \\ (\alpha + 1) \frac{\sum_{i \in \text{leaf}} \frac{\partial \mathcal{L}(y_i, z)}{\partial z}}{\sum_{i \in \text{leaf}} \frac{\partial^2 \mathcal{L}(y_i, z)}{\partial z^2}} \end{aligned}$$

MSE

The gradient we have as

$$\frac{\partial \mathcal{L}(y_i, z)}{\partial z}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}} = -2 \left(y_i - \frac{\alpha o_i + \gamma_0}{\alpha + 1} \right)$$

And hessian we have as

$$\frac{\partial^2 \mathcal{L}(y_i, z)}{\partial z^2}\Big|_{z = \frac{\alpha o_i + \gamma_0}{\alpha + 1}} = 2$$

So that

$$
\begin{aligned}
-\frac{G}{H} = \\
\sum_{i=1}^{n}(\alpha+1)y_i - \alpha o_i - \gamma_0 = \\
\sum_{i=1}^{n}(\alpha+1)y_i - \sum_{i=1}^{n}\alpha o_i - \sum_{i=1}^{n}\sum_{i=1}^{n}\frac{y_i}{n} = \\
\alpha\sum_{i=1}^{n}(y_i - o_i)
\end{aligned}
$$