Adaptive Weights Math

Let our ensemble be a weighted sum of trees:

$$\hat{y}(x) = \sum_{j=1}^{M} w_j f_j(x)$$

We wish to minimize an overall loss

$$L(w) = \sum_{i=1}^{N} l\left(y_i, \sum_{j=1}^{M} w_j f_j(x_i)\right)$$

Subject to $||w||_1 = \sum_{j=1}^{M} w_j = 1$. We assume the loss is twice differentiable.

Consider some small update $w^* = w^{(t)} + \Delta w$. We then have by Taylor Expansion that

$$L(w^{(t)} + \Delta w) = L(w^{(t)}) + g^T \Delta w + \frac{1}{2} \Delta w^T H \Delta w$$

Where $g = \nabla_w L(w^{(t)})$ and $H = \nabla_w^2 L(w^{(t)})$. We wish to minimize this loss, subject to the constraint of $\mathbf{1}^T \Delta w = 0$. This is equivalent to the original problem.

We thus have a Lagrangian formulation in

$$L(\Delta w, \lambda) = g^T \Delta w + \frac{1}{2} \Delta w^T H \Delta w + \lambda \mathbf{1}^T \Delta w$$

Since $\frac{\partial L(\Delta w, \lambda)}{\partial \Delta w} = 0$, we have that

$$g + H \Delta w + \lambda \mathbf{1} = 0$$

Supposing that H is invertible, we have that

$$\Delta w = -H^{-1}(g + \lambda \mathbf{1})$$

Working with the constraint, we have that

$$\begin{aligned}
\mathbf{1}^T \Delta w = 0 &\implies \\
\mathbf{1}^T(-H^{-1}(g + \lambda \mathbf{1})) = 0 &\implies \\
\lambda \mathbf{1}^T H^{-1} \mathbf{1} = -\mathbf{1}^T H^{-1} g &\implies \\
\lambda = -\frac{\mathbf{1}^T H^{-1} g}{\mathbf{1}^T H^{-1} \mathbf{1}} &
\end{aligned}$$

Thus, the update

$$\Delta w = -H^{-1} \left(g - \frac{\mathbf{1}^T H^{-1} g}{\mathbf{1}^T H^{-1} \mathbf{1}} \mathbf{1} \right)$$