$F = \{f_1, f_2, f_3, f_4 - - - - f_m\}$ set of base learners

Final prediction: $\hat{y}_i = \sum_{t=1}^{m} f_t(x_i)$

$D = \{x_1, x_2, x_3 - - - - x_n\}$ data points

$$L^{<t>} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{<t-1>} + f_t(x_i)) + \Omega(f_t) \text{ ——— } ①$$

Taylor Expansion

$$f(a+h) = f(a) + f'(a)h + \frac{1}{2} f''(a) h^2 + - - - - - f^n(a) \frac{h^n}{n!}$$

Here $a = \hat{y}_i^{<t-1>}$

$h = f_t(x_i)$

$f(a) = l(y_i, \hat{y}_i^{<t-1>})$

$$\therefore L^{<t>} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{<t-1>}) + \left(\frac{\partial l(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>}}\right) f_t(x_i) + \left(\frac{\partial^2 l(y_i, \hat{y}_i^{<t-1>})}{\partial \hat{y}_i^{<t-1>2}}\right) f_t(x_i)^2 +$$

$l(y_i, \hat{y}_i^{<t-1>})$ is constant ~~irr~~ irrespective of any function

$$\therefore L^{<t>} = \sum_{i=1}^{n} (C + g_i f_t(x_i) + h_i f_t(x_i)) + \Omega(f_t)$$

pick $f_t(x_i) \ni L^{<t>}$ is minimum. Removing constant as it is equal for any function.

$$L^{<t>} = \sum_{i=1}^{n} (g_i f_t(x_i) + h_i f_t(x_i)) + \Omega(f_t) \longrightarrow ②$$

Let f_t ~~(x)~~ has K leaf nodes. I_j be the set of instances belonging to node 'j'. 'W_j' be the prediction for node 'j'.

$$\Omega(f_t) = \gamma K + \frac{1}{2} \lambda \sum_{J=1}^{K} W_j^2$$

$$L^{<t>} = \sum_{j=1}^{K} \left[\left(\sum_{i \in I_j} g_i\right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda\right) w_j^2 \right] + \gamma K \longrightarrow ③$$

For each leaf 'j', $\frac{dL^{<t>}}{dw_j^*} = 0$

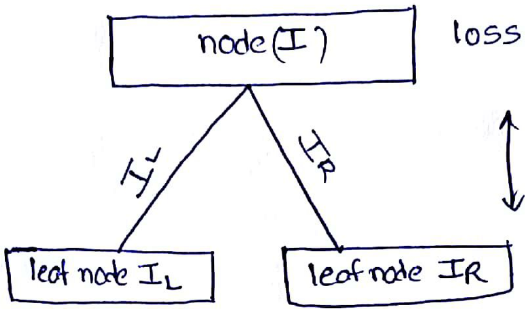
$$0 = \sum_{I \in I_j} g_i + \frac{1}{\cancel{2}} \left(\sum_{i \in I_j} h_i + \lambda\right) \cancel{2} \times w_j^*$$

$$\boxed{w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}}$$

Substituting weights into ③

$$\boxed{L^{<t>} = -\frac{1}{2}\sum_{j=1}^{K} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma K} \quad \text{——— ④}$$

This is the best loss for a fixed base learner with 'K' nodes. There will be several hundreds of possible tree structures. It is impossible to explore all of them.



loss according to ③ $\displaystyle -\frac{1}{2}\frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} + \gamma(1)$

reduction in loss

$$-\frac{1}{2}\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \gamma(1) \qquad -\frac{1}{2}\frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} + \gamma(1)$$

$$L_{split} = -\frac{1}{2}\frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} - \left[-\frac{1}{2}\left(\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda}\right)\right] + \gamma - \gamma - \gamma$$

$$= \boxed{\frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma}$$