2a. We want to solve for $\hat{\gamma}_{jm} = \underset{\gamma_{jm}}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma_{jm})$

Derive γ_{jm} for MSE:

Let $F(\gamma_{jm}) = \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma_{jm})$

$$= \sum_{x_i \in R_{jm}} \left[y_i - (f_{m-1}(x_i) + \gamma_{jm})\right]^2$$

$$= \sum_{x_i \in R_{jm}} \left[y_i^2 - 2y_i(f_{m-1}(x_i) + \gamma_{jm}) + (f_{m-1}(x_i) + \gamma_{jm})^2\right]$$

$$= \sum_{x_i \in R_{jm}} y_i^2 - 2y_i f_{m-1}(x_i) - 2y_i\gamma_{jm} + f_{m-1}(x_i)^2 + 2f_{m-1}(x_i)\gamma_{jm} + \gamma_{jm}^2$$

$$\therefore \frac{\partial F(\gamma_{jm})}{\partial \gamma_{jm}} = \sum_{x_i \in R_{jm}} -2y_i + 2f_{m-1}(x_i) + 2\gamma_{jm}$$

Setting $\frac{\partial F(\hat{\gamma}_{jm})}{\partial \hat{\gamma}_{jm}} = 0$, we obtain $\sum_{x_i \in R_{jm}} \hat{\gamma}_{jm} = \sum_{x_i \in R_{jm}} y_i - f_{m-1}(x_i)$

$$n_j \hat{\gamma}_{jm} = \sum_{x_i \in R_{jm}} y_i - f_{m-1}(x_i)$$

$$\hat{\gamma}_{jm} = \frac{1}{n_j} \sum_{x_i \in R_{jm}} y_i - f_{m-1}(x_i)$$

# of samples in terminal node j.

Derive γ_{jm} for binomial deviance loss:

Let $F(\gamma_{jm}) = \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma_{jm})$

$$= \sum_{x_i \in R_{jm}} -\log\left(1 + e^{-2y_i(f_{m-1}(x_i) + \gamma_{jm})}\right)$$

$$\therefore \frac{\partial F(\gamma_{jm})}{\partial \gamma_{jm}} = \sum_{x_i \in R_{jm}} \frac{-e^{-2y_i(f_{m-1}(x_i) + \gamma_{jm})}}{1 + e^{-2y_i(f_{m-1}(x_i) + \gamma_{jm})}} \cdot -2y_i$$

Setting $\frac{\partial F(\hat{\gamma}_{jm})}{\partial \hat{\gamma}_{jm}} = 0$, we obtain $\sum_{x_i \in R_{jm}} y_i \cdot e^{-2y_i(f_{m-1}(x_i) + \gamma_{jm})} = 0.$

This is hard to solve, so we would need to use second order approximation to minimize loss. We'll do so in 2b.

2b. Let $F(\gamma_{jm}) = \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma_{jm})$

To minmize F, we start by expressing L using second order approximation:

$$L(y_i, f_{m-1}(x_i)+\gamma_{jm}) \approx L(y_i, f_{m-1}(x_i)) + \frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i))\gamma_{jm} + \frac{1}{2}\frac{\partial^2 L(y_i, f_{m-1}(x_i))}{\partial f^2}\cdot \gamma_{jm}^2$$

By setting $\frac{\partial F(\hat{\gamma}_{jm})}{\partial \hat{\gamma}_{jm}} = \frac{\partial \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i)+\hat{\gamma}_{jm})}{\partial \hat{\gamma}_{jm}} = 0$,

We have:

$$\frac{\partial}{\partial \hat{\gamma}_{jm}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i)) + \frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i))\hat{\gamma}_{jm} + \frac{1}{2}\frac{\partial^2 L(y_i, f_{m-1}(x_i))}{\partial f^2}\hat{\gamma}_{jm}^2 = 0.$$

$$\sum_{x_i \in R_{jm}} \frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i)) + \frac{\partial^2 L(y_i, f_{m-1}(x_i))}{\partial f^2}\hat{\gamma}_{jm} = 0$$

$$\hat{\gamma}_{jm} = \frac{-\sum_{x_i \in R_{jm}} \frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i))}{\sum_{x_i \in R_{jm}} \frac{\partial^2}{\partial f^2} L(y_i, f_{m-1}(x_i))}.$$

For MSE:

$$L(y_i, f_{m-1}(x_i)) = (y_i - f_{m-1}(x_i))^2$$

$$\Rightarrow -\frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i)) = 2(y_i - f_{m-1}(x_i))$$

$$\Rightarrow \frac{\partial^2}{\partial f^2} L(y_i, f_{m-1}(x_i)) = \frac{\partial}{\partial f} 2(f_{m-1}(x_i) - y_i) = 2.$$

$$\therefore \hat{\gamma}_{jm} = \frac{\sum_{x_i \in R_{jm}} 2(y_i - f_{m-1}(x_i))}{\sum_{x_i \in R_{jm}} 2} = \frac{1}{n_j} \sum_{x_i \in R_{jm}} y_i - f_{m-1}(x_i)$$

# of samples in terminal node j.

For binomial deviance loss:

$$L(y_i, f_{m-1}(x_i)) = -(y_i f_{m-1}(x_i) - \log(1+e^{f_{m-1}(x_i)}))$$

$$\Rightarrow -\frac{\partial}{\partial f} L(y_i, f_{m-1}(x_i)) = \frac{\partial}{\partial f} y_i \cdot f_{m-1}(x_i) - \log(1+e^{f_{m-1}(x_i)})$$

$$= y_i - \frac{e^{f_{m-1}(x_i)}}{1+e^{f_{m-1}(x_i)}}$$

$$\Rightarrow \frac{\partial^2}{\partial f^2} L(y_i, f_{m-1}(x_i)) = \frac{\partial}{\partial f} \frac{e^{f_{m-1}(x_i)}}{1+e^{f_{m-1}(x_i)}} - y_i$$

$$= e^{f_{m-1}(x_i)}\left(1+e^{f_{m-1}(x_i)}\right)^{-1} - e^{2f_{m-1}(x_i)}\left(1+e^{f_{m-1}(x_i)}\right)^{-2}$$

By letting $p(x_i) = \frac{e^{f_{m-1}(x_i)}}{1+e^{f_{m-1}(x_i)}}$, we obtain

$$\hat{\gamma}_{jm} = \frac{\sum_{x_i \in R_{jm}} y_i - p(x_i)}{\sum_{x_i \in R_{jm}} p(x_i) - p(x_i)^2}$$