4. Derive expression (10.12) of ESL for the update of β in AdaBoost

10.11 in the text shows that

$$(e^{\beta} - e^{-\beta}) \sum_{i=1}^{N} w_i^{(m)} + e^{\beta} \cdot \sum_{y_i \neq G(x_i)} w_i^{(m)}$$

We are given in (10.9) that:

$$(\beta_m, G_m) = \arg\min_{\beta, G} \sum_{i=1}^{N} w_i^{(m)} \exp(-\beta y_i G(x_i))$$

$$G_m = \arg\min_{G} \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(x_i))$$

We take partial derivatives and set to 0:

$$\sum_{i=1}^{N} w_i^{(m)} y_i G(x_i) \exp(-\beta y_i G(x_i)) = 0$$

Separate:

$$\sum_{y_i = G(x_i)} w_i^{(m)} \exp(-\beta) - \sum_{y_i \neq G(x_i)} w_i^{(m)} \exp(\beta) = 0$$

$$\sum_{y_i = G(x_i)} w_i^{(m)} \exp(-\beta) = \sum_{y_i \neq G(x_i)} w_i^{(m)} \exp(\beta)$$

Multiplying Both sides by e^{β} and through a series of steps:

$$\exp(2\beta) = \frac{\sum_{y_i = G(x_i)} w_i^{(m)}}{\sum_{y_i \neq G(x_i)} w_i^{(m)}}$$

$$= \frac{1 - err_m}{err_m}$$

*We define err_m as the minimized weighted error rate

$$err_m = \frac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i^{(m)}}$$

We arrive at 10.12: $\beta = \frac{1}{2} \log \frac{1 - err_m}{err_m}$ ▨