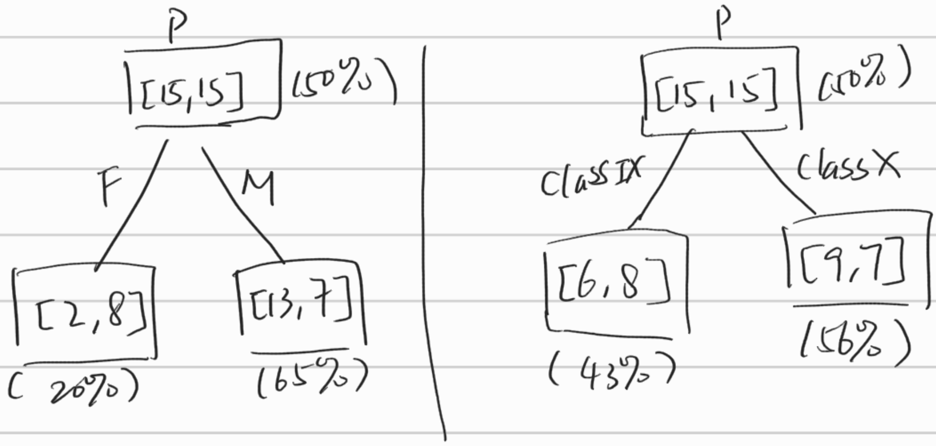
Q1 Split on gender



a) For "split on gender" find cross-entropy for parent node & each leaf

Cross entropy: $Q_m(T) = \sum_{k=1}^{K} P_{Tk} \ln P_{Tk}$

Parent cross entropy $= \frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2} = -0.693$ //

F leaf node cross entropy: $\frac{2}{10}\ln\frac{2}{10} + \frac{8}{10}\ln\frac{8}{10} = -0.5$ //

M leaf node cross entropy: $\frac{7}{20}\ln\frac{7}{20} + \frac{13}{20}\ln\frac{13}{20} = -0.647$ //

b) Find IG of the splitting.

IG = Information Gain = | Parent cross-entropy − weighted sum of children cross entropy |

weight of female leaf: $\frac{10}{30}$; weight of male leaf: $\frac{20}{30}$

IG for the split on gender:

$|\text{parent cross entropy} - (W_F \cdot CE_F + W_m \cdot CE_M)|$

$= |-0.693 - (\frac{10}{30}(-0.5) + \frac{20}{30}(-0.647))|$

$= |-0.693 - (-0.167 - 0.431)| = |-0.693 - (-0.598)| = 0.095$ //

c) Repeat a & b for "split on class"

parent cross entropy: shown in a $= -0.693$

Class IX cross entropy: $0.43 \ln 0.43 + 0.57 \ln 0.57 = -0.683$

Class X cross entropy: $0.56 \ln 0.56 + 0.44 \ln 0.44 = -0.686$

Weight of Class IX $= \frac{14}{30}$ Weight of class X $= \frac{16}{30}$

IG for the split on class:

$|-0.693 - (\frac{14}{30}(-0.683) + \frac{16}{30}(-0.686))| = 0.008$ //

d) Split on gender is better as its IG is bigger. It means more reduction of impurity. //

Q2. Show that in AdaBoost, α_m is updated by

assumed $=1$

$$\alpha_m = \eta \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) \quad , \text{by } \frac{dE}{d\alpha_m} = 0$$

where $E = e^{-\alpha_m/2} \sum_{n \in T_m} w_n^{(m)} + e^{\alpha_m/2} \sum_{n \in M_m} w_n^{(m)}$ — version 1 of E

$$= \left(e^{\alpha_m/2} - e^{-\alpha_m/2}\right) \underbrace{\sum_{n=1}^{N} w_n^{(m)} \cdot I(y_m(x_n) \neq t_n)}_{\text{call it } C} + e^{-\alpha_m/2} \sum_{n=1}^{N} w_n^{(m)}$$ — version 2 of E

where T_m are data points correctly classified by $y_m(x)$

M_m are those misclassified points

Find $\frac{dE}{d\alpha_m}$, using version 2 of given equation of E

$$\frac{dE}{d\alpha_m} = \frac{d}{d\alpha_m}\left(e^{\alpha_m/2} - e^{-\alpha_m/2}\right) C + e^{-\alpha_m/2} \sum_{n=1}^{N} w_n^{(m)}$$

$$= \left[\frac{\alpha_m}{2} e^{\alpha_m/2} - \left(-\frac{\alpha_m}{2}\right) e^{-\alpha_m/2}\right] C + \left(-\frac{\alpha_m}{2}\right) e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$$= \frac{\alpha_m}{2}\left(e^{\alpha_m/2} + e^{-\alpha_m/2}\right) \sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n) - \frac{\alpha_m}{2} e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$$\Downarrow$$

set derivative $=0$ move negative term to RHS $\rightarrow$

$$\cancel{\frac{\alpha_m}{2}}\left(e^{\alpha_m/2} + e^{-\alpha_m/2}\right) \sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n) = \cancel{\frac{\alpha_m}{2}} e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$$\left(e^{\alpha_m/2} + e^{-\alpha_m/2}\right) \underbrace{\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n)}_{=C} {}^{☆} = e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$\Downarrow$ divide both sides by $\sum_{n=1}^{N} w_n^{(m)}$

☆Note:

$$\varepsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}} = \frac{C}{\sum_{n=1}^{N} w_n^{(m)}}$$

$$\left(e^{\alpha_m/2} + e^{-\alpha_m/2}\right) \varepsilon_m = e^{-\frac{\alpha_m}{2}}$$

$$e^{\alpha_m/2} \varepsilon_m = e^{-\frac{\alpha_m}{2}} - e^{-\frac{\alpha_m}{2}} \varepsilon_m$$

$$e^{\alpha_m/2} \varepsilon_m = e^{-\frac{\alpha_m}{2}} (1 - \varepsilon_m)$$

divide both sides by $e^{-\frac{\alpha_m}{2}}$

$$e^{\alpha_m} = \frac{1-\varepsilon_m}{\varepsilon_m}$$

both side take ln

$$\therefore \alpha_m = \ln\left(\frac{1-\varepsilon_m}{\varepsilon_m}\right) \;//$$