CSC411 Homework 2

Deadline: Wednesday, Oct. 3, at 11:59pm.

Name: Baijiang, Yuan

Student Number: 1002175335

***Question* 1**:

$\boldsymbol{a}$) $We\ have\ H(x) = \sum_x p(x) log_2\left(\frac{1}{p(x)}\right)$

$\because 0 \leq p(x) \leq 1$

$\therefore \frac{1}{p(x)} \geq 1$

$\therefore log_2\left(\frac{1}{p(x)}\right) > 0$

$\therefore p(x) log_2\left(\frac{1}{p(x)}\right) \geq 0$

$\because x \in \mathcal{X}\ , \mathcal{X}\ might\ be\ \{1,2,\dots,N\}$

$\therefore \sum_x p(x) log_2\left(\frac{1}{p(x)}\right) \geq 0$

$\therefore H(x)\ \geq 0$

$\therefore The\ entropy\ H(x)\ is\ not\ negative.$

$\boldsymbol{b}$) $\because\ \mathrm{KL}(p||q) = \sum_x p(x) log_2\left(\frac{p(x)}{q(x)}\right)\ , and\ p(x) > 0, q(x) > 0\ for\ all\ x.$

$\therefore\ \mathrm{KL}(p||q) = \sum_x p(x)\left(-log_2\left(\frac{q(x)}{p(x)}\right)\right)$

$\therefore\ \mathrm{KL}(p||q) \geq\ -log_2 \sum_x p(x)\ \frac{q(x)}{p(x)}$ $\qquad\qquad \#\ By\ Jensen's\ Inequality$

$\therefore\ \mathrm{KL}(p||q) =\ -log_2 \sum_x p(x)$

$\therefore\ \mathrm{KL}(p||q) = -log_2\ 1$

$\therefore\ \mathrm{KL}(p||q) = 0$

$\therefore \mathrm{KL}(p||q) is\ not\ negative.$

$\boldsymbol{c}) \because I(Y;X) = H(Y) - H(Y|X)$

$\because\ H(Y) = \sum_y p(y) log_2 \left(\frac{1}{p(y)}\right) = -\sum_y p(y) log_2 p(y)$

$\because\ H(Y|X) = -\sum_x \sum_y p(x,y) log_2 p(y|x)$

$\therefore\ I(Y;X) = -\sum_y p(y) log_2 p(y) - \left(-\sum_x \sum_y p(x,y) log_2 p(y|x)\right)$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y) log_2 p(y|x) - \sum_y p(y) log_2 p(y)$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y) log_2 p(y|x) - \sum_y \sum_x p(y,x) log_2 p(y)$ $\quad \# \ Use\ p(x) = \sum_y p(x,y)$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y) log_2 p(y|x) - p(y,x) log_2 p(y)$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y) log_2 \frac{p(x,y)}{p(x)} - p(y,x) log_2 p(y)$ $\quad \# Use\ p(y|x) = \frac{p(x,y)}{p(x)}$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y)\ (log_2 p(x,y) - log_2(x) - log_2(y))$

$\therefore\ I(Y;X) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$

$\therefore\ I(Y;X) = \mathrm{KL(p(x,y)||p(x)p(y))}$

$\boldsymbol{Question\ 2}$:

$we\ have\ , L(y,t) = \frac{1}{2}(y-t)^2\ and\ \bar{h}(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(x)$

$\because \frac{1}{m}\sum_{i=1}^{m} L(h_i(x),t)$

$= \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2}(h_i(x) - t)^2$

$= \frac{1}{m}\sum_{i=1}^{m} \frac{1}{2}(h_i(x)^2 - 2h_i(x)\cdot t + t^2)$

$= \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2$

$\because\ \frac{1}{m}\sum_{i=1}^{m} h_i(x)^2 = E[h(x)^2]\ and\ [\frac{1}{m}\sum_{i=1}^{m} h_i(x)\ ]^2 = (E[h(x)])^2$

$\because x^2\ is\ convex\ function\ , since\ x^p\ is\ convex\ when\ p \geq 1\ or\ p \leq 0$

$\therefore by\ Jensen's\ inequality\ defination\ , we\ have\ E[h(x)^2] \geq (E[h(x)])^2$

$\therefore \frac{1}{m}\sum_{i=1}^{m} h_i(x)^2 \geq [\frac{1}{m}\sum_{i=1}^{m} h_i(x)\ ]^2\ , then\ we\ ues\ this\ to\ have\ the\ following.$

$\therefore \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2 \geq [\frac{1}{2m}\sum_{i=1}^{m} h_i(x)\ ]^{\mathbf{2}} - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2$

$\therefore \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2 \geq \frac{1}{2}\ \bar{h}(x)^2 - \bar{h}(x)\cdot t + \frac{1}{2}t^2$

$\therefore \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2 \geq \frac{1}{2}(\ \bar{h}(x)^2 - 2\bar{h}(x)\cdot t + t^2)$

$\therefore \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2 \geq \frac{1}{2}(\bar{h}(x) - t)^2$

$\therefore \frac{1}{2m}\sum_{i=1}^{m} h_i(x)^2 - \frac{1}{m}\sum_{i=1}^{m} h_i(x)\ \cdot t + \frac{1}{2}t^2 \geq L(\bar{h}(x),t)$

$\therefore \frac{1}{m}\sum_{i=1}^{m} L(h_i(x),t) \geq L(\bar{h}(x),t)$

$\therefore we\ conclude\ that$ *the loss of the average estimator is smaller than the average loss of the estimators*

***Question* 3**:

$we\ have\ err_t = \frac{\sum_{i \in E} W_i}{\sum_{i \in 1}^{N} W_i}\ \ and\ 1 - err_t = \frac{\sum_{i \in E^c} W_i}{\sum_{i \in 1}^{N} W_i}$

$$\therefore \alpha_t = \frac{1}{2}\log_e \frac{1-err_t}{err_t} = \frac{1}{2}\log_e \frac{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i}$$

$\because w_i \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\}, in\ this\ case\ \ (h_t(x^{(i)}),\ t^{(i)})\ can\ only\ be\ (1,-1)\ or\ (-1,1).$

$\therefore\ t^{(i)} h_t(x^{(i)}) = -1\ when\ h_t(x^{(i)}) \neq t^{(i)}$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \exp\,(-\alpha_t\, t^{(i)} h_t(x^{(i)}))$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \exp\left(-\frac{1}{2}\log_e \frac{1-err_t}{err_t}\ (-1)\right)$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \exp\left(\frac{1}{2}\log_e \frac{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i)}\right)$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \left(e^{\frac{1}{2}\log_e \frac{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i)}} \right)$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \left(e^{\log_e \frac{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i)}} \right)^{\frac{1}{2}}$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} = w_i \left(\frac{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i)}\right)^{\frac{1}{2}} \qquad \#\ ①$$

$We\ using\ the\ same\ process\ above\ to\ find\ w_i' \mathbb{I}\{h_t(x^{(i)}) = t^{(i)}\}, the\ only\ difference\ is$ $(h_t(x^{(i)}),\ t^{(i)})\ is\ equal\ to (1,1)\ or\ (-1,-1) this\ time,$

$\therefore\ t^{(i)} h_t(x^{(i)}) = 1\ when\ h_t(x^{(i)}) = t^{(i)}$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) = t^{(i)}\} = w_i \exp\left(-\frac{1}{2}\log_e \frac{1-err_t}{err_t}\ (1)\right)$$

$$\therefore w_i' \mathbb{I}\{h_t(x^{(i)}) = t^{(i)}\} = w_i \left(\frac{\sum_{i \in E} W_i \,/\, \sum_{i \in 1}^{N} W_i)}{\sum_{i \in E^c} W_i \,/\, \sum_{i \in 1}^{N} W_i}\right)^{\frac{1}{2}} \qquad \#\ ②$$

$$\because w_i' = w_i' \mathbb{I}\{h_t(x^{(i)}) = t^{(i)}\} + w_i' \mathbb{I}\{h_t(x^{(i)}) \neq t^{(i)}\} \qquad \#\ ③=①+②$$

$From ①, ②\ and\ ③. Now\ we\ can\ solve\ ett_r'$

$$\because We\ know:\ err_t' = \frac{\sum_{i=1}^{N} w_i' \mathbb{1}\{h_t(x^{(i)}) \neq t^{(i)}\}}{\sum_{i=1}^{N} w_i'}$$

$$\therefore err_t' = \frac{\sum_{i=1}^{N} w_i' \mathbb{1}\{h_t(x^{(i)}) \neq t^{(i)}\}}{\sum_{i=1}^{N} w_i' \mathbb{1}\{h_t(x^{(i)}) \neq t^{(i)}\} + \sum_{i=1}^{N} w_i' \mathbb{1}\{h_t(x^{(i)}) = t^{(i)}\}} \qquad \#Using\ ③$$

$$\therefore err_t' = \frac{\sum_{i \in E} w_i \left(\frac{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}\right)^{\frac{1}{2}}}{\sum_{i \in E} w_i \left(\frac{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}\right)^{\frac{1}{2}} + \sum_{i \in E^c} w_i \left(\frac{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}\right)^{\frac{1}{2}}} \qquad \#Using\ ①and②$$

$$\therefore err_t' = \frac{\sqrt{(\sum_{i \in E} w_i)^2} \cdot \sqrt{\frac{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}}}{\sqrt{((\sum_{i \in E} w_i))^2} \cdot \sqrt{\frac{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}} + \sqrt{(\sum_{i \in E^c} w_i)^2} \cdot \sqrt{\frac{\sum_{i \in E} w_i / \sum_{i \in 1}^{N} w_i)}{\sum_{i \in E^c} w_i / \sum_{i \in 1}^{N} w_i}}}$$

$$\therefore err_t' = \frac{\sqrt{(\sum_{i \in E} w_i)^2 \cdot \frac{\sum_{i \in E^c} w_i}{\sum_{i \in E} w_i}}}{\sqrt{(\sum_{i \in E} w_i)^2 \cdot \frac{\sum_{i \in E^c} w_i}{\sum_{i \in E} w_i}} + \sqrt{(\sum_{i \in E^c} w_i)^2 \cdot \frac{\sum_{i \in E} w_i}{\sum_{i \in E^c} w_i}}}$$

$$\therefore err_t' == \frac{\sqrt{\sum_{i \in E} w_i \cdot \sum_{i \in E^c} w_i}}{\sqrt{\sum_{i \in E} w_i \cdot \sum_{i \in E^c} w_i} + \sqrt{\sum_{i \in E^c} w_i \cdot \sum_{i \in E} w_i}}$$

$$\therefore err_t' = \frac{1}{2}$$

Interpretation: From AdaBoost, we know that at each iteration we re-weight the training samples by assigning larger weights to samples that were classified incorrectly. Then we will train a new weak classifier based on the re-weighted samples. $\mathrm{err}_t' = \frac{1}{2}$ indicates the error or each new classifier is at most 0.5.