Q1 a) $H(X) = \sum p(x) \log(\frac{1}{p(x)})$

$= \sum p(x) (-\log p(x))$

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

$\therefore \log p(x) \le 0$

$\therefore -\log p(x) \ge 0.$

$\therefore -p(x) \log p(x) \ge 0$

$\therefore \sum p(x)(-\log(p(x))) \ge 0.$

$\therefore \sum p(x) \log(\frac{1}{p(x)}) \ge 0.$

Q1 b) $KL(p||q) = \sum p(x) \log_2 \frac{p(x)}{q(x)}$

$-KL(p||q) = \sum p(x) \log \frac{q(x)}{p(x)}$

$= \sum p(x) \log \frac{q(x)}{p(x)}$

$\because \log(\frac{\sum x_i}{n}) \ge \frac{\sum(\log x_i)}{n}.$

(by Jensen's Inequality)

$\le \log \sum p(x) \frac{q(x)}{p(x)}$

$= \log \sum q(x)$

$= \log 1$

$= 0.$

$\therefore KL(p||q) \ge 0.$

c) $I(Y;X) = KL(p(x,y) || p(x)p(y))$.

$RHS = I(Y;X)$

$= H(Y) - H(Y|X).$

$= \sum_y p(y) \log_2 \frac{1}{p(y)} - \sum p(x) H(Y|X=x)$

$= \sum_y p(y) \log_2 \frac{1}{p(y)} - \cancel{\sum p(x)} \sum p(x,y) \log \frac{p(x)}{p(x,y)}$

$L.H.S. = KL(p(x,y) || p(x)p(y)) = \sum p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$

$= -\sum (x,y) \log_2 \frac{p(x)p(y)}{p(x,y)}.$

$= -\sum p(x,y) \log \frac{p(x)}{p(x,y)} - \sum p(x,y) \log p(y).$

$= -\sum p(x,y) \log \frac{p(x)}{p(x,y)} + \sum p(x,y) \log \frac{1}{p(y)}$

$LHS = RHS.$

Q2.

$$\bar{h}(x) = \frac{1}{m} \sum h_i(x).$$

$$L(\bar{h}(x), t) \leq \frac{1}{m} \sum L(h_i(x), t).$$

$$LHS = L(\tfrac{1}{m}\sum h(x_i), t)$$
$$= L(\Sigma(h(x_i)), t)$$

$$RHS = E(h(x_i), t)$$

$$L(y, t) = \frac{1}{2}(y - t)^2$$

$\because \ p = 2 \geq 1.$

$\therefore$ $L(y,t)$ is convex.

$\therefore \ L(\bar{h}(x), t) \leq \frac{1}{m} \sum L(h_i(x), t)$

Q3.

$$err' = \frac{\sum^N w' I\{h_t(x) \neq t^i\}}{\sum^N w'}$$

$$= \frac{\sum^{z} w e^{\wedge}(-\alpha t^i h_t(x)) \overbrace{I\{h_t(x) \neq t_i\}}^{=1} + \sum^{z^c} w e^{\wedge}(-\alpha t^i h_t(x))}{\sum^{z} w e^{\wedge}(2\alpha \underbrace{I\{h_t(x) \neq t_i\}}_{=1}) + \sum^{z^c} w \underbrace{e^{\wedge}(2\alpha I\{h_t(x_i) \neq t_i\})}_{=1}} = 0.$$

$$= \frac{\sum^{z} w_i e^{\wedge}(2\alpha I\{h_t(x) \neq t_i\})}{\sum^{z} w e^{\wedge}(2\alpha) \sum^{z^c} w}$$

$$= \frac{\frac{\sum^{z} w_i}{\sum w_i}}{\frac{\sum^{z} w_i}{\sum^N w_i} + \frac{\sum^{z^c} w \cdot e^{\wedge}(-2\alpha)}{\sum^N w}}$$

$$= \frac{\text{error}}{\text{error} + e^{\wedge}(2\alpha)(1 - \text{error})}$$

$$= \frac{\text{error}}{\text{error} + \frac{\text{error}}{1-\text{error}}(1 - \text{error})}$$

$$= \frac{1}{2}.$$

Q3 b) Interpretation of the result:
The purpose of boosting is to combine many different hypothesis from hypothesis space so that we end up with better final hypothesis. The error rate of 1/2 means high diversity and less probability of overfitting. The error rate is 0.5 such that the alpha will be 0 in the next iteration. error is the same for future iteration.