1.a). $H(x) = \sum_x p(x) \log_2 \frac{1}{p(x)}$

This will be non-negative if both terms to be multiplied are positive, therefore the addition of positive terms will be positive.

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

$$1/0 > 1/p(x) \geq 1$$

$$\infty > 1/p(x) \geq 1$$

$$\log_2 1 \leq \log_2 1/p(x) < \log_2 \infty$$

$$\left[0 \leq \log_2 1/p(x) < \infty \right] \times p(x)$$

$$\sum_x \left[0 \leq p(x) \log_2 1/p(x) < \infty \right]$$

$$0 \leq \sum_x p(x) \log_2 1/p(x) < \infty$$

$0 \leq H(x) < \infty \rightarrow \therefore H(x)$ is ≥ 0, being non-negative.

QED.

1.6). $p(x) > 0$, $q(x) > 0$

$$KL(p\|q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}$$

Let $\log_2 \frac{p(x)}{q(x)} = f(x)$

$$\Rightarrow KL(p\|q) = \sum_x p(x) f(x)$$
$$= \underset{p}{E}(f(x))$$

$\Rightarrow$ $f(x)$ is concave since, according to the Appendix, it is a log function whose values, $\frac{p(x)}{q(x)} > 0$ are positive real numbers.

$$\Rightarrow \left.\begin{array}{ll} \phi(E(x)) \le E[\phi(x)] & \text{if convex} \\ \phi[E(x)] \ge E[\phi(x)] & \text{if concave} \end{array}\right] \to \text{Jensen's Ineq.}$$

$\Rightarrow$ Taking the above properties, we can say the following...

$$-KL(p\|q) = -\underset{p}{E}[f(x)] = \sum_x p(x) \log_2 \frac{q(x)}{p(x)}$$
$$\le \log_2 E_p\left(\frac{q(x)}{p(x)}\right)$$
$$= \log_2 \sum_x \cancel{p(x)} \left(\frac{q(x)}{\cancel{p(x)}}\right)$$
$$= \log_2 \underbrace{\sum_x q(x)}_{=1}$$
$$= \log_2 1$$
$$= 0$$

$$\Rightarrow \left(-KL(p\|q) \le 0\right) \times -1$$

$KL(p\|q) \ge 0$ $\therefore$ $KL(p\|q)$ is non-negative.

QED

1. c). $IG(Y,X) = H(Y) - H(Y|X)$

$= -H(Y|X) + H(Y)$ Taken from lecture slides.

$= \sum_{x,y} p(x,y) \log_2 p(y|x) - \sum_{y} p(y) \log_2 p(y)$

Marginal dist. of y

$= \sum_{x,y} p(x,y) \log_2 p(y|x) - \sum_{y} \left[\sum_{x} p(x,y) \right] \log_2 p(y)$

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

$= \sum_{x,y} \left[p(x,y) \log_2 \frac{p(x,y)}{p(x)} - p(x,y) \log_2 p(y) \right]$

$= \sum_{x,y} p(x,y) \left[\log_2 \frac{p(x,y)}{p(x)} - \log_2 p(y) \right]$

$= \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$ (1)

=> We know that $KL(p||q) = \sum_{x} p(x) \log_2 \frac{p(x)}{q(x)}$.

Extending to two dimensions, we can say that

$KL(p||q) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{q(x,y)}$ (2)

=> In the same way, taking equation 1, we can see that it is in a similar form as equation 2. Taking that example, we can say...

$IG(Y,X) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} = KL(p(x,y) || p(x)p(y))$ as desired.

QED.

2. $L(\bar{h}(x), t)$

$= \frac{1}{2}(\bar{h}(x) - t)^2$

$= \frac{1}{2}\Big(\underbrace{\frac{1}{m}\sum_{1}^{m} h_i(x)}_{= E[h(x)]} - t\Big)^2$

$= \frac{1}{2}\big(E[h(x)] - t\big)^2$

$\Rightarrow$ Let us say that $E[h(x)] - t = u$

Then $L(\bar{h}(x), t) = \frac{1}{2}(u)^2$ which is convex by the Appendix.

This means that $f(E[h(x)], t) = L(\bar{h}(x), t)$ is convex.

So $f(E[h(x)], t) \leq E[f(h(x), t)]$ by Jensen's Ineq.

$\Rightarrow L(\bar{h}(x), t) = \frac{1}{2}(E[h(x)] - t)^2 \leq E[\frac{1}{2}(h(x) - t)^2]$

$= E[L(h(x), t)]$

$= \frac{1}{m}\sum_{1}^{m} L(h_i(x), t)$ as desired.

QED.

3. $$err'_t = \frac{\sum_1^N w_i' I\{h_t(x^i) \neq t^i\}}{\sum_1^N w_i'}$$

$$= \frac{\sum_1^N w_i e^{-\alpha_t t^i h_t(x^i)} I\{h_t(x^i) \neq t^i\}}{\sum_1^N w_i e^{-\alpha_t t^i h_t(x^i)}}$$

=> The only values of i are those that $h_t(x^i) = t^i$ or $h_t(x^i) \neq t^i$. We can split the sums as such.

=> Note: $I\{h_t(x^i) \neq t^i\} = \begin{Bmatrix} 1, & h_t(x^i) \neq t^i \\ 0, & h_t(x^i) = t^i \end{Bmatrix}$

$$\Rightarrow err_t = \frac{\sum_1^N w_i I\{h_t(x^i) \neq t^i\}}{\underbrace{\sum_1^N w_i}_{=1}}$$

$$= \sum_1^N w_i I\{h_t(x^i) \neq t^i\}$$

$$= \sum_{\substack{i \to h_t(x^i) \neq t^i \\ \in N}} w_i \underbrace{I\{h_t(x^i) \neq t^i\}}_{1} + \sum_{i \in N \to h_t(x^i) = t^i} w_i \underbrace{I\{h_t(x^i) = t^i\}}_{0}$$

$$= \sum_{i \in N \to h_t(x^i) \neq t^i} w_i$$

$$\Rightarrow err'_t = \frac{\sum_{i \in N \to h_t(x^i) \neq t^i} w_i e^{-\alpha_t \overbrace{[t^i h_t(x^i)]}^{= -1(1) \text{ or } (1)(-1) \; = -1}} \overbrace{I\{h_t(x^i) \neq t^i\}}^{1} + \sum_{i \in N \to h_t(x^i) = t^i} w_i e^{-\alpha_t t^i h_t(x^i)} \overbrace{I\{h_t(x^i) = t^i\}}^{0}}{\sum_{i \in N \to h_t(x^i) \neq t^i} w_i e^{-\alpha_t \underbrace{[t^i h_t(x^i)]}_{= -1}} + \sum_{i \in N \to h_t(x^i) = t^i} w_i e^{-\alpha_t \underbrace{(t^i h_t(x^i))}_{= 1(1) \text{ or } -1(-1) \; = 1}}}$$

$$= \frac{\sum_{i \in N \to h_t(x^i) \neq t^i} w_i e^{+\alpha_t}}{\sum_{i \in N \to h_t(x^i) \neq t^i} w_i e^{+\alpha_t} + \sum_{i \in N \to h_t(x^i) = t^i} w_i e^{-\alpha_t}}$$

$$\Rightarrow \sum_{i \in N \to h_t(x^i) \neq t^i} w_i = err_t$$

$$\therefore 1 - err_t = \sum_{i \in N \to h_t(x^i) = t^i} w_i$$

$$= \left[\frac{e^{\alpha_t} err_t}{e^{\alpha_t} err_t + e^{-\alpha_t}(1 - err_t)} \right] \times \frac{e^{-\alpha_t}}{e^{-\alpha_t}}$$

$$= \frac{err_t}{err_t + e^{-2\alpha_t}(1 - err_t)}$$

$\Rightarrow$ We know $\alpha_t = \frac{1}{2} \ln \frac{1 - err_t}{err_t}$

$$-\alpha_t = \frac{1}{2} \ln \frac{err_t}{1 - err_t}$$

$$-2\alpha_t = \ln \frac{err_t}{1 - err_t}$$

$$e^{-2\alpha_t} = \frac{err_t}{1 - err_t}$$

$$\Rightarrow err'_t = \frac{err_t}{err_t + e^{-2\alpha_t}(1 - err_t)} = \frac{err_t}{err_t + \frac{err_t}{\cancel{(1 - err_t)}} \cancel{(1 - err_t)}} = \frac{err_t}{err_t + err_t}$$

$$= \frac{\cancel{err_t}}{2 \cancel{err_t}}$$

$$= \boxed{1/2}$$

QED