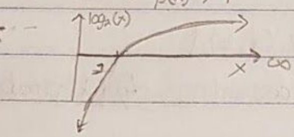
CSC 411 HW2 Jan 28, 2017

1. a) Axiom of probability: $0 \le p(x) \le 1$

$\Rightarrow \infty \ge \frac{1}{p(x)} \ge 1$

$\log_2(x)$:



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

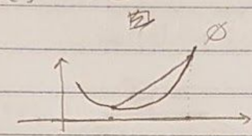
$\therefore \log_2\left(\frac{1}{p(x)}\right) \ge 0$

$\therefore p(x) \log_2 \frac{1}{p(x)} \ge 0$

$\therefore \sum_x p(x) \log_2 \frac{1}{p(x)} \ge 0$

$\boxtimes$

b) Jensen's Ineq:



$\phi(E[x]) \le E[\phi(x)]$

First note that not necessarily each term ≥ 0. e.g. $\frac{1/4}{1/2} = \frac{p(x)}{q(x)}$

$\phi(x) = \log(x)$ is concave. $\Rightarrow E[\phi(x)] \le \phi(E[x])$

Let $\phi(x) = -\log(x)$

$$-\sum_x p(x) \log_2 \frac{p(x)}{q(x)} = \sum p(x) \log_2 \frac{q(x)}{p(x)} \le \log_2\Big(\underbrace{\sum_x p(x) \frac{q(x)}{p(x)}}_{=1 \text{ by prob. axioms}}\Big) = 0$$

$\therefore$ Since $-KL(p\|q) \le 0$, $KL(p\|q) \ge 0$

$\boxtimes$

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

$KL(p(x,y) \| p(x)p(y)) = \sum_x \sum_y p(x,y) \log_2\left(\frac{p(x,y)}{p(x)p(y)}\right)$

Recall: $p(y) = \sum_x p(x,y)$

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

$= \sum_x \sum_y p(x,y) \log_2 \frac{p(y|x)}{p(y)}$

$p(y|x) = \frac{p(x,y)}{p(x)}$

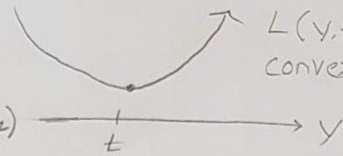
$\therefore H(Y) - H(Y|X) = KL(p(x,y) \| p(x)p(y))$

as required.

JE:

$\phi(E[x]) \le E(\phi(x))$

$f(\lambda x + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2)$



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

Show:

2. $\frac{1}{2}\left(\frac{1}{m}\sum_{i=1}^{m} h_i(x) - t\right)^2 \le \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{1}{2}(h_i(x) - t)^2$

$L(E[y], t) \le E[L(y,t)]$

$E[y] = \frac{1}{m}\sum_{i=1}^{m} h_i(x)$ assuming equal prob.

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

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

---

Proof of convexity: $\frac{dL}{dy} = (y-t)$

$\frac{d^2L}{dy^2} = 1 \geq 0 \quad \forall y$

$\therefore$ convex.

---

if $t^i = h_t(x^i)$

$$w_i' = w_i\sqrt{\frac{err_t}{1-err_t}}$$

if $t^i \neq h_t(x^i)$, $t^i h_t(x^i) = -1$

$$w_i' = w_i\left(\sqrt{\frac{1-err_t}{err_t}}\right)$$

3. $$w_i' \leftarrow w_i e^{-\alpha_t t^i h_t(x^i)} = w_i\left(e^{\log\frac{err_t}{1-err_t}}\right)^{\frac{1}{2}t^i h_t(x^i)} = w_i\left(\frac{err_t}{1-err_t}\right)^{\frac{1}{2}t^i h_t(x^i)}$$

$$err_t' = \frac{\sum w_i' \mathbb{I}\{h_t(x^i) \neq t^i\}}{\sum w_i'}$$

Let W be set of misclassified

Let C be set of correctly classified

$W + C = N$

$$= \frac{\left(\sum_W w_i\sqrt{\frac{1-err_t}{err_t}}\right)}{\left(\underbrace{\sum_W w_i\sqrt{\frac{1-err_t}{err_t}}}_{\text{weights for mis classified}} + \underbrace{\sum_C w_i\sqrt{\frac{err_t}{1-err_t}}}_{\text{weights for correctly classified}}\right)}$$

what is err_t? $err_t = \dfrac{\sum_W w_i}{\sum_W w_i + \sum_C w_i}$

$$err_t' = \frac{\sum_W w_i}{\sum_W w_i + \sum_C w_i \frac{err_t}{1-err_t}}$$ $\leftarrow$ mult top/bottom $\bar{w}\sqrt{\frac{err_t}{1-err_t}}$

$$\frac{err_t}{1-err_t} = \frac{\frac{\sum_W w_i}{\sum_W w_i + \sum_C w_i}}{1 - \frac{\sum_W w_i}{\sum_W w_i + \sum_C w_i}} = \frac{\frac{\sum_W w_i}{\sum_W w_i + \sum w_i}}{\frac{\sum_C w_i}{\sum_W w_i + \sum_C w_i}}$$

$$\therefore \frac{err_t}{1-err_t} = \frac{\sum_W w_i}{\sum_C w_i}$$

$$err_t' = \frac{\sum_W w_i}{\sum_W w_i + \cancel{\sum_C w_i}\frac{\sum_W w_i}{\cancel{\sum_C w_i}}}$$

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

QED

Interpretation:

- mathematically, weights are s.t. this previously optimal hypothesis is guareenteed to perform the worst given the new weights. Therefore, guaranteed to not continuously select the same hypothesis from $\mathcal{H}$.