A3Q2

2022年6月29日 星期三 下午4:47

a)

① When $t=1$, we initialize $p^t = [1/n, 1/n \cdots]$ and $\tilde{p}^t = [1/n, 1/n, -]$
so $p^0 = \tilde{p}^0$

② Hypothesis: Assume that for some $k \geq 1$, $p^k = \tilde{p}^k$

③ We need to prove that $p^{k+1} = \tilde{p}^{k+1}$

since we have $\tilde{y}_i = \frac{y_i+1}{2}$, $\tilde{h}(x_i) = \frac{h(x_i)+1}{2}$

then: $\tilde{y}_i = \tilde{h}(x_i) \iff \frac{y_i+1}{2} = \frac{h(x_i)+1}{2} \iff y_i = h(x_i)$

Also: $p^k = \tilde{p}^k$ by our assumption, so:

$$\epsilon_k = \sum_{i=1}^{n} p_i^k [h_k(x_i) \neq y_i] = \sum_{i=1}^{n} \tilde{p}_i^k [\tilde{h}_k(x_i) \neq \tilde{y}_i] = \tilde{\epsilon}_k$$

$$\beta_k = \frac{1}{2} \log \frac{1-\epsilon_k}{\epsilon_k} \qquad \tilde{\beta}_k = \frac{\epsilon_k}{1-\epsilon_k}$$

$$e^{\beta_k} = e^{\frac{1}{2}\log\frac{1-\epsilon_k}{\epsilon_k}} = \left(\frac{1-\epsilon_k}{\epsilon_k}\right)^{\frac{1}{2}} = \tilde{\beta}_k^{-\frac{1}{2}}$$

$$p_i^{k+1} = \frac{w_i^{k+1}}{\sum_{j=1}^n w_j^{k+1}} = \frac{w_i^k e^{-y_i \beta_k h_k(x_i)}}{\sum_{j=1}^n w_j^k e^{-y_j \beta_k h_k(x_j)}} = \frac{w_i^k \left(\tilde{\beta}_k^{-\frac{1}{2}}\right)^{-y_i h_k(x_i)}}{\sum_{j=1}^n w_j^k \left(\tilde{\beta}_k^{-\frac{1}{2}}\right)^{-y_j h_k(x_j)}}$$

$$= \frac{w_i^k (\tilde{\beta}_k)^{\frac{1}{2} y_i h_k(x_i)}}{\sum_{j=1, h_k(x_j)=y_j}^{n} w_j^k \tilde{\beta}_k^{\frac{1}{2}} + \sum_{j=1, h_k(x_j)\neq y_j}^{n} w_j^k \tilde{\beta}_k^{-\frac{1}{2}}} \quad \left(\begin{cases} y_i \cdot h_k(x_i) = 1 & y_i = h_k(x_i) \\ y_i h_k(x_i) = -1 & y_i \neq h_k(x_i) \end{cases} \right)$$

$$= \begin{cases} \dfrac{w_i^k}{\sum_{h_k(x_j)=y_j} w_j^k + \sum_{h_k(x_j)\neq y_j} w_j^k \tilde{\beta}_k^{-1}} & \text{when } h_k(x_i) = y_i \\[2ex] \dfrac{w_i^k}{\sum_{h_k(x_j)=y_j} w_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} w_j^k} & \text{when } h_k(x_i) \neq y_i \end{cases}$$

when $y_i = h_k(x_i)$, either $\begin{cases} y_i = 1 \\ h_k(x_i) = 1 \end{cases}$ or $\begin{cases} y_i = -1 \\ h_k(x_i) = -1 \end{cases} \Rightarrow h_k(x_i) y_i = 1$

when $y_i \neq h_k(x_i)$, either $\begin{cases} y_i = -1 \\ h_k(x_i) = 1 \end{cases}$ or $\begin{cases} y_i = 1 \\ h_k(x_i) = -1 \end{cases} \Rightarrow h_k(x_i) y_i = -1$

when $\tilde{y}_i = \tilde{h}_k(x_i)$, either $\begin{cases} \tilde{y}_i = 1 \\ h_k(x_i) = 1 \end{cases}$ or $\begin{cases} \tilde{y}_i = 0 \\ h_k(x_i) = 0 \end{cases} \Rightarrow |\tilde{h}_k(x_i) - \tilde{y}_i| = 0$

When $\tilde{y}_i = \tilde{h}_k(x_i)$, either $\begin{cases} \tilde{y}_i = 1 \\ h_k(x_i) = 1 \end{cases}$ or $\begin{cases} \tilde{y}_i = 0 \\ h_k(x_i) = 0 \end{cases}$ $\Rightarrow |\tilde{h}_k(x_i) - \tilde{y}_i| = 0$

When $\tilde{y}_i \neq \tilde{h}_k(x_i)$, either $\begin{cases} \tilde{y}_i = 0 \\ h_k(x_i) = 1 \end{cases}$ or $\begin{cases} \tilde{y}_i = 1 \\ h_k(x_i) = 0 \end{cases}$ $\Rightarrow |\tilde{h}_k(x_i) - \tilde{y}_i| = 1$

$$\tilde{P}_i^{k+1} = \frac{\tilde{W}_i^k \tilde{\beta}_k^{1-|\tilde{h}_k(x_i)-y_i|}}{\sum_{j=1}^{n} \tilde{W}_j^k \tilde{\beta}_k^{1-|\tilde{h}_k(x_j)-y_j|}} = \frac{\tilde{W}_i^k \tilde{\beta}_k^{1-|\tilde{h}_k(x_i)-y_i|}}{\sum_{h_k(x_j)=y_j} \tilde{W}_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} \tilde{W}_j^k \tilde{\beta}_k^0}$$

$$= \begin{cases} \dfrac{\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} \tilde{W}_j^k + \sum_{h_k(x_j)\neq y_j} \tilde{W}_j^k \tilde{\beta}_k^{-1}} & \text{when } h_k(x_i) = y_i \\ \dfrac{\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} \tilde{W}_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} \tilde{W}_j^k} & \text{when } h_k(x_i) \neq y_i \end{cases}$$

From our hypothesis: $P_i^t = \tilde{P}_i^t$, we have $\dfrac{W_i^t}{\sum_{j=1}^n W_j^t} = \dfrac{\tilde{W}_i^t}{\sum_{j=1}^n \tilde{W}_j^t}$ $\forall i$

so: $\dfrac{W_i^t}{\tilde{W}_i^t} = \dfrac{\sum_{j=1}^n W_j^t}{\sum_{j=1}^n \tilde{W}_j^t}$, a constant.

Let's use m to denote $\dfrac{W_i^t}{\tilde{W}_i^t} \Rightarrow W_i^t = m \tilde{W}_i^t$

Therefore:

$$P_i^{k+1} = \begin{cases} \dfrac{W_i^k}{\sum_{h_k(x_j)=y_j} W_j^k + \sum_{h_k(x_j)\neq y_j} W_j^k \tilde{\beta}_k^{-1}} & \text{when } h_k(x_i) = y_i \\ \dfrac{W_i^k}{\sum_{h_k(x_j)=y_j} W_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} W_j^k} & \text{when } h_k(x_i) \neq y_i \end{cases}$$

$$= \begin{cases} \dfrac{m\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} m\tilde{W}_j^k + \sum_{h_k(x_j)\neq y_j} m\tilde{W}_j^k \tilde{\beta}_k^{-1}} & \text{when } h_k(x_i) = y_i \\ \dfrac{m\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} mW_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} mW_j^k} & \text{when } h_k(x_i) \neq y_i \end{cases}$$

$$h_k(x_j)=y_j \qquad h_k(x_j)\neq y_j$$

$$= \begin{cases} \dfrac{\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} \tilde{W}_j^k + \sum_{h_k(x_j)\neq y_j} \tilde{W}_j^k \tilde{\beta}_k^{-1}} & \text{when } h_k(x_i)=y_i \\[2ex] \dfrac{\tilde{W}_i^k}{\sum_{h_k(x_j)=y_j} \tilde{W}_j^k \tilde{\beta}_k + \sum_{h_k(x_j)\neq y_j} \tilde{W}_j^k} & \text{when } h_k(x_i)\neq y_i \end{cases}$$

$$= \tilde{p}_i^{k+1}$$

Therefore: $p_i^k = \tilde{p}_i^k \Rightarrow p_i^{k+1} = \tilde{p}_i^{k+1}$

③ Conclusion, By the induction above, we have proved that $p_i^t = \tilde{p}_i^t \quad \forall\, i, t \quad i=1\dots n, \ t=1\dots T$.

As long as p_i^t are equal, ϵ_t will be equal and β_t will also be equal. (because $h(x_i)\neq y \Leftrightarrow \tilde{h}(x_i) = \tilde{y}_i$)

Therefore, the new updates is the same as the original one.

b) $E_{y\sim p}[\, e^{-yH} \mid X=x]$

$$= p(y=1|X=x)\cdot e^{-H} + p(y=-1|X=x)\, e^{H} = f$$

$$\frac{df}{dH} = -p(y=1|X=x)\, e^{-H} + p(y=-1|X=x)\, e^{H}$$

$$\frac{d^2f}{dH^2} = p(y=1|X=x)\, e^{H} + p(y=-1|X=x)\, e^{H} \geq 0$$

since $p\geq 0,\ e^{x}>0$

So f is convex, we can get its minimum at x where $\frac{df}{dH}(x)=0$

$$-p(y=1|X=x)\, e^{-H} + p(y=-1|X=x)\, e^{H} = 0$$

$$p(y=-1|X=x)\, e^{H} = p(y=1|X=x)\, e^{-H}$$

$$e^{2H} = \frac{p(y=1|X=x)}{p(y=-1|X=x)}$$

$$H = \frac{1}{2}\log\frac{p(y=1|X=x)}{p(y=-1|X=x)}$$

Therefore, the minimier H is proportional to the log odds ratio.

c) :

$$\bar{\epsilon}_t = \epsilon(\bar{h}_t) = \sum_{i=1}^{n} p_i^t [\bar{h}_t(x_i) \neq y_i]$$

$$= \sum_{i=1}^{n} p_i^t [h_t(x_i) = y_i]$$ since $h_t(x_i) \in \{0,1\}$, $-h_t(x_i) = \bar{h}_t(x_i)$

$$= 1 - \sum_{i=1}^{n} p_i^t [h_t(x_i) \neq y_i]$$ since $\sum_{i=1}^{n} p_i^t = 1$

$$= 1 - \epsilon_t$$

$$\bar{\beta}_t = \frac{1}{2}\log\frac{1-\bar{\epsilon}_t}{\bar{\epsilon}_t} = \frac{1}{2}\log\frac{1-(1-\epsilon_t)}{1-\epsilon_t} = \frac{1}{2}\log\frac{\epsilon_t}{1-\epsilon_t} = -\frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}$$

$$= -\beta_t$$

$$\bar{W}_i^{t+1} = \bar{W}_i^t \, e^{-y_i \bar{\beta}_t \bar{h}_t(x_i)}$$

$$= \bar{W}_i^t \, e^{-y_i(-\beta_t)\cdot(-h_t(x_i))} = \bar{W}_i^t \, e^{-y_i \beta_t h_t(x_i)}$$

We do get the same update in (4) if we are using the original β_t and h_t.

In other words, we will use the same formula after set h_t to $\bar{h}_t$ to update W, no matter whether we use $h_t(x)$ and β_t or $\bar{h}_t(x)$ and $\bar{\beta}_t$

d)

Set $g = \hat{E}_t \, e^{-y\beta h_t(x)}$, $(x,y) \sim p^t$.

$$g = \sum_{i=1}^{n} p_i^t \cdot e^{-y\beta h_t(x)}$$

$$= \sum_{h_t(x_i)=y_i} p_i^t e^{-\beta} + \sum_{h_t(x_i)\neq y_i} p_i^t e^{\beta}$$

Notice that p_i^t is a constant as $\beta_1 \cdots \beta_{t-1}$ should be determined by now.

$$\frac{dg}{d\beta} = -\sum_{h_t(x_i)=y_i} p_i^t e^{-\beta} + \sum_{h_t(x_i)\neq y_i} p_i^t e^{\beta}$$

$$\frac{d^2 g}{d\beta^2} = \sum_{h_t(x_i)=y_i} p_i^t e^{-\beta} + \sum_{h_t(x_i)\neq y_i} p_i^t e^{\beta} > 0$$ since $p_i^t \geq 0$, $e^x > 0$

so g is convex, the optimal β is when $\frac{dg}{d\beta} = 0$

$$\Leftrightarrow \sum_{h_t(x_i)\neq y_i} p_i^t e^{\beta} = \sum_{h_t(x_i)=y_i} p_i^t e^{-\beta}$$

$$\Leftrightarrow \quad \sum_{h_t(x_i)\neq y_i} p_i\, e = \sum_{h_t(x_i)=y_i} p_i\, e$$

$$\Leftrightarrow \quad e^{\beta} \sum_{h_t(x_i)\neq y_i} p_i^t = e^{-\beta} \sum_{h_t(x_i)=y_i} p_i^t$$

$$\Leftrightarrow \quad e^{2\beta} = \frac{\sum_{h_t(x_i)=y_i} p_i^t}{\sum_{h_t(x_i)\neq y_i} p_i^t} = \frac{\sum_{i=1}^{n} p_i^t\,[h_t(x_i)=y_i]}{\sum_{i=1}^{n} p_i^t\,[h_t(x_i)\neq y_i]}$$

since $\sum_{i=1}^{n} p_i^t = 1$

$$= \frac{1-\sum_{i=1}^{n} p_i^t\,[h_t(x_i)\neq y_i]}{\sum_{i=1}^{n} p_i^t\,[h_t(x_i)\neq y_i]} = \frac{1-\epsilon_t}{\epsilon_t}$$

$$\Leftrightarrow \quad \beta = \frac{1}{2}\log\frac{1-\epsilon_t}{\epsilon_t}.$$

e)

$$\epsilon_{t+1} = \sum_{i=1}^{n} p_i^{t+1}\,[h_t(x_i)\neq y_i] = \sum_{h_t(x_i)\neq y_i} p_i^{t+1}$$

$$= \sum_{h_t(x_i)\neq y_i} \frac{W_i^t}{\sum_{h_t(x_j)=y_j} W_j^t\,\tilde{\beta}_t + \sum_{h_t(x_j)\neq y_j} W_j^t} \qquad \text{from a)}$$

$$= \frac{1}{\sum_{h_t(x_j)=y_j} W_j^t \frac{\epsilon_t}{1-\epsilon_t} + \sum_{h_t(x_j)\neq y_j} W_j^t} \cdot \sum_{i=1}^{n} W_i^t\,[h_t(x_i)\neq y_i]$$

$$= \frac{\sum_{i=1}^{n} W_i^t\,[h_t(x_i)\neq y_i]}{\frac{\epsilon_t}{1-\epsilon_t}\sum_{j=1}^{n} W_j^t\,[h_t(x_j)=y_j] + \sum_{j=1}^{n} W_j^t\,[h_t(x_j)\neq y_j]}$$

$$= \frac{\epsilon_t}{\frac{\epsilon_t}{1-\epsilon_t}\cdot(1-\epsilon_t) + \epsilon_t} = \frac{\epsilon_t}{2\,\epsilon_t} = \frac{1}{2}.$$