HW # 6 CMSC 25025 Xinyu Wei

1. (a) If a point is correctly classified:

$$Y^t g(x) = \left(-\frac{1}{C-1}\right)^2 \cdot (C-1) + 1^2$$
$$= \frac{1}{C-1} + 1^2$$
$$= \frac{C}{C-1}$$

If a point is incorrectly classified:

$$Y^t g(x) = \left(-\frac{1}{C-1}\right)^2 \cdot (C-2) + \left(-\frac{1}{C-1}\right) \times 2$$
$$= \frac{C-2}{(C-1)^2} + \frac{-2}{C-1}$$
$$= \frac{C-2+2-2C}{(C-1)^2}$$
$$= \frac{-C}{(C-1)^2}$$

(b). $\tilde{u}_i = \exp\left[-\frac{1}{C}\left(\sum_{r=1}^{m-1} \beta_r Y_i^t g_r(X_i)\right)\right]$

For β fixed, $g_m = \arg\min_g \sum_{i=1}^n \exp\left[-\frac{1}{C}\left(\sum_{r=1}^{m-1} \beta Y_i^t g_r(X_i) + \beta Y_i^t g(X_i)\right)\right]$

$$= \arg\min_g \sum_{i=1}^n \exp\left[-\frac{1}{C}\left(\sum_{r=1}^{m-1} \beta Y_i^t g_r(X_i)\right)\right] \cdot \exp\left[-\frac{1}{C}\beta Y_i^t g(X_i)\right]$$
$$= \arg\min_g \sum_{i=1}^n \tilde{u}_i \cdot \exp\left[-\frac{1}{C}\beta Y_i^t g(X_i)\right]$$
$$= \arg\min_g \sum_{Y_i^t g(X_i)<0} \tilde{u}_i \exp\left[-\frac{1}{C}\beta \cdot \frac{-C}{(C-1)^2}\right] + \sum_{Y_i^t g(X_i)>0} \tilde{u}_i \exp\left[-\frac{1}{C}\beta \frac{C}{C-1}\right]$$

Let $A = \exp\left(\frac{\beta}{(C-1)^2}\right)$

$B = \exp\left(\frac{\beta}{1-C}\right) < 1$

$A > B$

we can see $A = B^{\frac{1}{1-C}}$

$$= \arg\min_g \sum_{i=1}^n \tilde{u}_i \left(\mathbb{I}[Y_i^t g(X_i)<0] \cdot A + \mathbb{I}[Y_i^t g(X_i)>0] \cdot B\right)$$
$$= \arg\min_g \sum_{i=1}^n u_i \left(\mathbb{I}[Y_i^t g(X_i)<0] \cdot A + \mathbb{I}[Y_i^t g(X_i)>0] \cdot B\right)$$

since $\sum_{i=1}^n \tilde{u}_i$ is fixed number dividing the equation by $\sum_{i=1}^n \tilde{u}_i$ doesn't change the minimization problem.

$$= \arg\min_g \sum_{i=1}^n u_i \left[\mathbb{I}[Y_i^t g(X_i)<0] \cdot A + (1-\mathbb{I}[Y_i^t g(X_i)>0]) \cdot B\right]$$
$$= \arg\min_g \sum_{i=1}^n u_i \left[(A-B) \cdot \mathbb{I}[Y_i^t g(X_i)<0] + B\right] \quad (*)$$

Since $A > B$, $A-B > 0$

$u_i > 0$, $B > 0$, $A-B$ and B are fixed.

So minimizing $(*)$ is to minimize $\sum_{i=1}^n u_i \mathbb{I}[Y_i^t g(X_i)<0] = e_m(g)$

Thus, $g_m = \arg\min_g e_m(g)$

Now g is fixed, $\sum_{i=1}^n u_i \cdot \mathbb{I}[Y_i^t g(X_i)<0] = e_m$

$$\beta_m = \arg\min_\beta \sum_{i=1}^n u_i \left[(A-B) \cdot \mathbb{I}[Y_i^t g(X_i)<0] + B\right]$$
$$= \arg\min_\beta \sum_{i=1}^n u_i \cdot \mathbb{I}(Y_i^t g(X_i)<0) \cdot (A-B) + \sum_{i=1}^n u_i \cdot B$$
$$= \arg\min_\beta e_m (A-B) + B$$

plug in $A = B^{\frac{1}{1-c}}$. $\hookrightarrow = \underset{\beta}{\arg\min}\ e_m(B^{\frac{1}{1-c}} - B) + B$

$= \underset{\beta}{\arg\min}\ (1-e_m)B + e_m \cdot B^{\frac{1}{1-c}}$ $(**)$

$$\frac{d(**)}{dB} = 1 - e_m + e_m \cdot \left(-\frac{1}{c-1}\right) \cdot B^{-\frac{1}{c-1}-1}$$

$$\frac{d(**)}{dB} = 0 \iff 1 - e_m = \frac{e_m}{c-1} B_*^{\frac{c}{1-c}}$$

$$\frac{(1-e_m)(c-1)}{e_m} = B_*^{\frac{c}{1-c}}$$

$$B_* = \left(\frac{(1-e_m)(c-1)}{e_m}\right)^{\frac{1-c}{c}}$$

So B_* minimize $(**)$

$$\beta_* = (1-c) \cdot \log B_*$$
$$= (1-c) \cdot \frac{1-c}{c} \cdot \left[\log \frac{(1-e_m)}{e_m} + \log(c-1)\right]$$
$$= \frac{(c-1)^2}{c} \cdot \left[\log \frac{(1-e_m)}{e_m} + \log(c-1)\right]$$

(c) $u = [u_1, \cdots, u_n]$

$w = [w_1, \cdots, w_n]$

u and w are both initialized uniformly.

So at 1st step, $u_i = w_i \quad \forall i = 1, 2, \cdots, n$

Now we show that if at k^{th} step, $u_i = w_i$, then at $(k+1)$th step, $u_i = w_i$

For any $i, j = 1, 2, \cdots, n$

$$\frac{u_i^{(k)}}{u_j^{(k)}} = \frac{w_i^{(k)}}{w_j^{(k)}} \quad \text{since } u_i = w_i,\ u_j = w_j \text{ at } k\text{th step.}$$

Now we discuss 4 cases:

① if $g(x_i)$ incorrect, $g(x_j)$ correct.

$$\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{w_i^{(k)}}{w_j^{(k)}} \cdot (c-1) \cdot \frac{1-e_m}{e_m}$$

$$\frac{u_i^{(k+1)}}{u_j^{(k+1)}} = \frac{u_i^{(k)}}{u_j^{(k)}} \cdot \frac{\exp\left[-\frac{(c-1)^2}{c^2}\left(\log\frac{1-e_m}{e_m} + \log(c-1)\right) \cdot \frac{-c}{(c-1)^2}\right]}{\exp\left[-\frac{(c-1)^2}{c^2}\left(\log\frac{1-e_m}{e_m} + \log(c-1)\right) \cdot \frac{c}{c-1}\right]}$$

$$= \frac{u_i^{(k)}}{u_j^{(k)}} \cdot \exp\left[\left(\frac{1}{c} + \frac{c-1}{c}\right) \cdot \left(\log\frac{1-e_m}{e_m} + \log(c-1)\right)\right]$$

$$= \frac{u_i^{(k)}}{u_j^{(k)}} \cdot \exp\left[\log\frac{(1-e_m)}{e_m} + \log(c-1)\right] = \frac{u_i^{(k)}}{u_j^{(k)}} \cdot (c-1) \cdot \frac{1-e_m}{e_m} = \frac{w_i^{(k)}}{w_j^{(k)}} \cdot (c-1) \cdot \frac{1-e_m}{e_m} = \frac{w_i^{(k+1)}}{w_j^{(k+1)}}$$

So $\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{u_i^{(k+1)}}{u_j^{(k+1)}}$ for any i, j s.t. $g(x_i)$ is incorrect
$g(x_j)$ is correct. (α)

② $g(x_i)$ is correct, $g(x_j)$ is incorrect

Similar to proof in case ① since we could simply have $\frac{w_j^{(k+1)}}{w_i^{(k+1)}} = \frac{u_j^{(k+1)}}{u_i^{(k+1)}}$ from (α)

③ $g(x_i)$ and $g(x_j)$ are both incorrect

$$\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{w_i^{(k)}}{w_j^{(k)}} \cdot \frac{(C-1)\frac{1-e_m}{e_m}}{(C-1)\frac{1-e_m}{e_m}} = \frac{w_i^{(k)}}{w_j^{(k)}}$$

$$\frac{u_i^{(k+1)}}{u_j^{(k+1)}} = \frac{u_i^{(k)}}{u_j^{(k)}} \cdot \frac{\exp[-\frac{(C-1)^2}{C}(\log\frac{1-e_m}{e_m} + \log(C-1)) \cdot \frac{-C}{(C-1)^2}]}{\exp[-\frac{(C-1)^2}{C}(\log\frac{1-e_m}{e_m} + \log(C-1)) \cdot \frac{-C}{(C-1)^2}]}$$

$$= \frac{u_i^{(k)}}{u_j^{(k)}} \cdot 1$$

$$= \frac{w_i^{(k)}}{w_j^{(k)}} = \frac{w_i^{(k+1)}}{w_j^{(k+1)}}$$

So $\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{u_i^{(k+1)}}{u_j^{(k+1)}}$ for any i, j s.t. $g(x_i)$ and $g(x_j)$ are both incorrect.

④ $g(x_i)$ and $g(x_j)$ are both correct

Similar to case ③, we can get $\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{w_i^{(k)}}{w_j^{(k)}} = \frac{u_i^{(k)}}{u_j^{(k)}} = \frac{u_i^{(k+1)}}{u_j^{(k+1)}}$

So $\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{u_i^{(k+1)}}{u_j^{(k+1)}}$ for any i, j s.t. $g(x_i)$ and $g(x_j)$ are both correct

Denotation: $w_i^{(k+1)}, w_j^{(k+1)}$ are w_i and w_j at $(k+1)$th step
$w_i^{(k)}, w_j^{(k)}$ are w_i and w_j at kth step
Similarly for $u_i^{(k+1)}, u_j^{(k+1)}, u_i^{(k)}, u_j^{(k)}$

From ①②③④, we've shown that $\frac{w_i^{(k+1)}}{w_j^{(k+1)}} = \frac{u_i^{(k+1)}}{u_j^{(k+1)}}$ for any $i, j = 1, 2, \cdots, n$.

And since $\sum_{i=1}^{n} u_i^{(k+1)} = 1$, $\sum_{i=1}^{n} w_i^{(k+1)} = 1$,

we must have that $u_i^{(k+1)} = w_i^{(k+1)}$ for any $i = 1, 2, \cdots, n$.

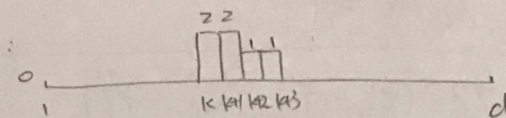
Thus, if $w_i = u_i$ at kth step, $w_i = u_i$ at $(k+1)$th step.

In conclusion, $u_i = w_i$ $\forall i = 1, 2, \cdots, n$ at all the step.

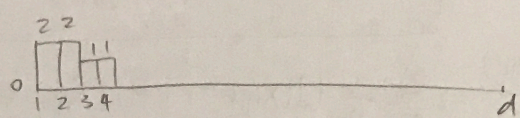
3. (a). If $Y=k$

X will look like $[0, 0, \cdots 0, 2, 2, 1, 1, 0, \cdots 0]$ with positions 1, $k-1$, k, $k+1$, $k+2$, $k+3$, d marked by arrows

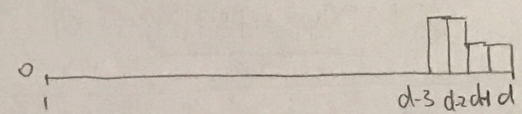
geometrically :



if $Y=1$: X will look like



if $Y=d-3$: X will look like



Geometrically, Y represents the start of signal of X.

(b). $\log P(X_1, \cdots, X_d, Y)$

$= \log\left(P(X_1, \cdots, X_d \mid Y) \cdot P(Y) \right)$

$= \log\left(P(X_1 \mid Y) \cdot P(X_2 \mid Y) \cdots P(X_d \mid Y) \cdot P(Y) \right)$

$= \log\left(a_{X_Y} \cdot a_{X_{Y+1}} \cdot b_{X_{Y+2}} \cdot b_{X_{Y+3}} \cdot \prod_{\substack{j<Y \\ \text{or } j \geq Y+4}} C_{X_j} \cdot \pi_Y \right)$

$= \log a_{X_Y} + \log a_{X_{Y+1}} + \log b_{X_{Y+2}} + \log b_{X_{Y+3}} + \log \pi_Y + \sum_{\substack{j<Y \text{ or} \\ j \geq Y+4}} \log C_{X_j}$

(c). Fully Observed data X_i, Y_i, $i = 1, 2, \cdots, n$

$\log P(X, Y \mid a, b, c, \pi_k)$

$= \log \prod_{i=1}^{n} P(X_i, Y_i \mid a, b, c, \pi_k)$

$= \log \prod_{i=1}^{n} P(X_{i1}, X_{i2}, \cdots, X_{id}, Y_i \mid a, b, c, \pi_k)$

$= \sum_{i=1}^{n} \log P(X_{i1}, X_{i2}, \cdots, X_{id}, Y_i \mid a, b, c, \pi_k)$

$= \sum_{i=1}^{n} \left(\sum_{j=1}^{d} \log P(X_{ij} \mid Y_i, a, b, c, \pi_k) + \log \pi_{Y_i} \right)$

$= \sum_{i=1}^{n} \left(\log a_{X_{iY_i}} + \log a_{X_{iY_i+1}} + \log b_{X_{iY_i+2}} + \log b_{X_{iY_i+3}} + \sum_{\substack{j<Y_i \\ \text{or } j \geq Y_i+4}} \log C_{X_{ij}} + \log \pi_{Y_i} \right)$

$\hat{a}_r = \arg\max_{a_r} \log P(X, Y \mid a, b, c, \pi_k)$ for fixed b, c, π_k, subject to $a_0 + a_1 + a_2 = 1$

$\hat{b}_r = \arg\max_{b_r} \log P(X, Y \mid a, b, c, \pi_k)$ for fixed a, c, π_k, subject to $b_0 + b_1 + b_2 = 1$

$\hat{c}_r = \arg\max_{c_r} \log P(X, Y \mid a, b, c, \pi_k)$ for fixed a, b, π_k, subject to $c_0 + c_1 + c_2 = 1$

$\hat{\pi}_k = \arg\max_{\pi_k} \log P(X, Y \mid a, b, c, \pi_k)$ for fixed a, b, c, subject to $\sum_{k=1}^{d-3} \pi_k = 1$

$$\frac{d \log P(X,Y|a,b,c,\pi_k)}{d a_r} = \sum_{i=1}^{n} \cdot \frac{1}{a_r} \cdot (I[X_{iY_i}=r] + I[X_{iY_i+1}=r]) \qquad r=0,1,2$$

$$\frac{d \log P(X,Y|a,b,c,\pi_k)}{d a_r} = 0, \quad \sum_{r=0}^{2} a_r = 1$$

$$\Rightarrow \text{ Solved as } \boxed{\hat{a}_r = \frac{\sum_{i=1}^{n}(I[X_{iY_i}=r] + I[X_{iY_i+1}=r])}{2n}, \quad r=0,1,2}$$

$$\frac{d \log P(X,Y|a,b,c,\pi_k)}{d b_r} = \sum_{i=1}^{n} \frac{1}{b_r} (I[X_{iY_i+2}=r] + I[X_{iY_i+3}=r]) = 0, \quad \sum_{r=0}^{2} b_r = 1$$

$$\text{Solved as } \boxed{\hat{b}_r = \frac{\sum_{i=1}^{n}(I[X_{iY_i+2}=r] + I[X_{iY_i+3}=r])}{2n}, \quad r=0,1,2}$$

$$\frac{d \log P(X,Y|a,b,c,\pi_k)}{d c_r} = \sum_{i=1}^{n} \frac{1}{c_r} \cdot \Big(\sum_{\substack{j<Y_i \\ \text{or } j>Y_i+4}} I[X_{ij}=r]\Big) = 0, \quad \sum_{r=0}^{2} c_r = 1$$

$$\text{Solved as } \boxed{\hat{c}_r = \frac{\sum_{i=1}^{n}\Big(\sum_{\substack{j<Y_i \\ \text{or } j>Y_i+4}} I[X_{ij}=r]\Big)}{(d-4)\,n}, \quad r=0,1,2}$$

$$\frac{d \log P(X,Y|a,b,c,\pi_k)}{d \pi_k} = \sum_{i=1}^{n} \frac{1}{\pi_k} \cdot I[Y_i=k] \qquad k=1,2,\cdots,d-3$$

$$= 0 \quad \text{and} \quad \sum_{k=1}^{d-3} \pi_k = 1$$

$$\text{Solved as } \boxed{\hat{\pi}_k = \frac{\sum_{i=1}^{n} I(Y_i=k)}{n}, \quad k=1,2,\cdots,d-3}$$

(d). Y_i's unobserved

① E step: $\sum_{i=1}^{n} E_{Y_i|X_i,\theta^{old}} l(X_i,Y_i|\theta)$

$$= \sum_{i=1}^{n} \Big[\sum_{Y_i} P(Y_i|X_i,\theta^{old}) \cdot \log P(X_i,Y_i|\theta)\Big] \qquad (*)$$

For $P(Y_i|X_i,\theta^{old}) = \frac{P(X_i,Y_i|\theta)}{P(X_i|\theta)} = \frac{P(X_i,Y_i|\theta)}{\sum_{Y_i} P(X_i,Y_i|\theta)}$ let $v_{ik} = \log(P(Y_i=k, X_i|a,b,c,\pi))$

$$= \frac{\exp(v_{ik})}{\sum_{k=1}^{d-3}\exp(v_{ik})} \qquad \text{let } d = \log\Big(\sum_{k=1}^{d-3}\exp(v_{ik})\Big)$$

$$= \frac{\exp(v_{ik})}{\exp(d)}$$

$$= \exp(v_{ik}-d) \qquad \text{let } \exp(v_{ik}-d) = w_{ik}, \quad \sum_{k=1}^{d-3} w_{ik} = 1$$

So $(*) = \sum_{i=1}^{n} \left(\sum_{k=1}^{d-3} w_{ik} \cdot P(X_i, Y_i = k \mid \theta) \right)$

$$= \sum_{i=1}^{n} \sum_{k=1}^{d-3} w_{ik} \Big(\log a_{X_{ik}} + \log a_{X_{ik+1}} + \log b_{X_{ik+2}} + \log b_{X_{ik+3}} + \log \pi_k + \sum_{\substack{j<k \\ \text{or } j \geq k+4}} \log c_{x_{ij}} \Big)$$

②: M step: $\hat{a}, \hat{b}, \hat{c}, \hat{\pi} = \underset{a,b,c,\pi}{\text{argmax}} \ (*)$

Similar to what we did in part (c),

we get $\hat{a}_r = \dfrac{\sum_{i=1}^{n} \sum_{k=1}^{d-3} w_{ik} (I[X_{ik} = r] + I[X_{ik+1} = r])}{2n}$

$$\hat{b}_r = \frac{\sum_{i=1}^{n} \sum_{k=1}^{d-3} w_{ik} (I[X_{ik+2} = r] + I[X_{ik+3} = r])}{2n}$$

$$\hat{c}_r = \frac{\sum_{i=1}^{n} \sum_{k=1}^{d-3} w_{ik} \Big(\sum_{\substack{j<k \\ \text{or } j \geq k+4}} I[X_{ij} = r] \Big)}{(d-4) \cdot n} \qquad r = 0, 1, 2$$

$$\hat{\pi}_k = \frac{\sum_{i=1}^{n} w_{ik}}{n} \qquad k = 1, 2, \dots, d-3$$

psedo code: ⓪ Initialize a, b, c, π

① Compute w_{ik} given X, $a^{old}, b^{old}, c^{old}, \pi^{old}$

② update $a^{new} = \hat{a}_r$

$b^{new} = \hat{b}_r$

$c^{new} = \hat{c}_r$

$\pi^{new} = \hat{\pi}_k$

using w_{ik} in step ①

③. repeat ①,② until $\theta = \{a, b, c, \pi\}$,

$\theta^{new} - \theta^{old} < \text{tolerance}$