Assignment 3

Q1 (a) $P(X, C|\theta, \pi) = \prod_{i=1}^{N} P(X^{(i)}, C^{(i)})$

$$P(x, c|\theta, \pi) = \prod_{i=1}^{N}\left[P(C^{(i)}|\pi)\prod_{j=1}^{784} P(X_j^{(i)}|C^{(i)}, \theta_{jc^{(i)}})\right]$$

$$l(x,c|\theta,\pi) = \log P(x,c|\theta,\pi) = \sum_{i=1}^{N}\log\left[P(C^{(i)}|\pi)\prod_{j=1}^{784}P(X_j^{(i)}|C^{(i)},\theta_{jc^{(i)}})\right]$$

$$= \underbrace{\sum_{i=1}^{N}\log P(C^{(i)}|\pi)}_{A} + \underbrace{\sum_{j=1}^{784}\sum_{i=1}^{N}\log P(X_j^{(i)}|C^{(i)},\theta_{jc^{(i)}})}_{B}$$

To maximize $l(x,c|\theta,\pi)$, we maximize A and B individually

$$\frac{\partial l(x,c|\theta,\pi)}{\partial\theta} = \sum_{j=1}^{784}\sum_{i=1}^{N}\frac{\partial \log P(X_j^{(i)}|C^{(i)},\theta_{jc^{(i)}})}{\partial\theta}$$

θ_{jc}: probability of obtaining $X_j=1$ from class c

$$= \sum_{j=1}^{784}\sum_{i=1}^{N}\frac{\partial\log\left[\theta_{jc^{(i)}}^{x_j^{(i)}}(1-\theta_{jc^{(i)}})^{(1-x_j^{(i)})}\right]}{\partial\theta}$$

$$= \sum_{j=1}^{784}\sum_{i=1}^{N}\frac{\partial\left[X_j^{(i)}\log\theta_{jc^{(i)}} + (1-X_j^{(i)})\log(1-\theta_{jc^{(i)}})\right]}{\partial\theta}$$

$$= \sum_{j=1}^{784}\sum_{i=1}^{N}\left[X_j^{(i)}\frac{\partial\log\theta_{jc^{(i)}}}{\partial\theta} + (1-X_j^{(i)})\frac{\partial(1-\theta_{jc^{(i)}})}{\partial\theta}\right]$$

$$= \sum_{j=1}^{784}\sum_{i=1}^{N}X_j^{(i)}\frac{\partial\log\theta_{jc^{(i)}}}{\partial\theta} + \sum_{j=1}^{784}\sum_{i=1}^{N}(1-x_j^{(i)})\frac{\partial(1-\theta_{jc^{(i)}})}{\partial\theta}$$

$$= M \in \mathbb{R}^{784\times10} \quad \text{where } M_{jc} = \frac{\sum_{i=1}^{N}\mathbb{I}[X_j^{(i)}=1 \,\&\, C^{(i)}=c]}{\theta_{jc}} - \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, C^{(i)}=c]}{1-\theta_{jc}}$$

$$\frac{\partial l(x,c|\theta,\pi)}{\partial\theta} = 0 \iff M = 0$$

$$\iff M_{jc} = 0 \quad \forall j, c$$

$$\iff \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, C^{(i)}=c]}{\theta_{jc}} - \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, C^{(i)}=c]}{1-\theta_{jc}} = 0$$

$$(1-\theta_{jc})\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c] = \theta_{jc}\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]$$

$$\frac{1}{\theta_{jc}} - 1 = \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c] \Big/ \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]$$

$$\frac{1}{\theta_{jc}} = 1 + \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c] \Big/ \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]$$

$$\frac{1}{\theta_{jc}} = \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c] + \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]}$$

$$\theta_{jc} = \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]}{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c] + \sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}$$

$$\boxed{\theta_{jc} = \frac{\sum_{i=1}^{N}\mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]}{\sum_{i=1}^{N}\mathbb{I}[c^{(i)}=c]} = \frac{\#\text{ of sample where } j^{th} \text{ bit is 1 and class is } c}{\#\text{ of sample of class } c}}$$

$$\log P(t^{(i)}|\pi) = \log \prod_{j=0}^{9} \pi_j^{t_j^{(i)}}$$

$$= \sum_{j=0}^{9} \log \pi_j^{t_j^{(i)}}$$

$$= \sum_{j=0}^{8} \log \pi_j^{t_j^{(i)}} + \log \pi_9^{t_9^{(i)}}$$

$$= \sum_{j=0}^{8} t_j^{(i)} \log \pi_j + t_9^{(i)} \log\left(1 - \sum_{j=0}^{8}\pi_j\right)$$

$$\frac{\partial \log P(t^{(i)}|\pi)}{\partial \pi_k} = t_k^{(i)}\frac{1}{\pi_k} + t_9^{(i)}\frac{1}{1-\sum_{j=0}^{8}\pi_j}\cdot(-1) \quad \forall k \in \{0,1,\dots,8\}$$

$$= \frac{t_k^{(i)}}{\pi_k} - \frac{t_9^{(i)}}{\pi_9}$$

Therefore $\dfrac{\partial \ell(x, c|\theta,\pi)}{\partial \pi} = 0 \iff \displaystyle\sum_{i=1}^{N}\frac{\partial \log P(t^{(i)}|\pi)}{\partial \pi} = 0$

$$\iff \sum_{i=1}^{N}\frac{\partial \log P(t^{(i)}|\pi)}{\partial \pi_k} = 0 \quad \forall k \in \{0,1,\dots 9\}$$

$$\iff \sum_{i=1}^{N}\left[\frac{t_k^{(i)}}{\pi_k} - \frac{t_9^{(i)}}{\pi_9}\right] = 0$$

$$\frac{1}{\pi_k}\sum_{i=1}^{N} t_k^{(i)} = \frac{1}{\pi_9}\sum_{i=1}^{N} t_9^{(i)}$$

$$\pi_k = \pi_9 \cdot \sum_{i=1}^{N} t_k^{(i)} \Big/ \sum_{i=1}^{N} t_9^{(i)} \quad \forall k \in \{0, \dots, 8\}$$

We know $\sum_{k=0}^{8} \pi_k + \pi_9 = 1$

$$\sum_{k=0}^{8} \left(\pi_9 \cdot \sum_{i=1}^{N} t_k^{(i)} \Big/ \sum_{i=1}^{N} t_9^{(i)}\right) + \pi_9 = 1$$

$$\left(1 + \frac{\sum_{k=0}^{8}\sum_{i=1}^{N} t_k^{(i)}}{\sum_{i=1}^{N} t_9^{(i)}}\right) \cdot \pi_9 = 1$$

$$\pi_9 = \frac{\sum_{i=1}^{N} t_9^{(i)}}{\sum_{k=0}^{8}\sum_{i=1}^{N} t_k^{(i)} + \sum_{i=1}^{N} t_9^{(i)}}$$

$$\pi_9 = \frac{\sum_{i=1}^{N} t_9^{(i)}}{\sum_{k=0}^{9}\sum_{i=1}^{N} t_k^{(i)}} = \frac{\sum_{i=1}^{N} t_9^{(i)}}{N} = \frac{\sum_{i=1}^{N} \mathbb{I}[t_9^{(i)} = 1]}{N}$$

AND $\pi_k = \pi_9 \cdot \frac{\sum_{i=1}^{N} t_k^{(i)}}{\sum_{i=1}^{N} t_9^{(i)}}$

$$= \frac{\sum_{i=1}^{N} t_9^{(i)}}{N} \cdot \frac{\sum_{i=1}^{N} t_k^{(i)}}{\sum_{i=1}^{N} t_9^{(i)}}$$

$$= \frac{\sum_{i=1}^{N} t_k^{(i)}}{N} = \frac{\sum_{i=1}^{N} \mathbb{I}[t_k^{(i)} = 1]}{N}$$

Therefore $\boxed{\pi_k = \frac{\sum_{i=1}^{N} \mathbb{I}[t_k^{(i)} = 1]}{N} = \frac{\#\text{ of sample of class } k}{\#\text{ of sample}} \quad \text{for } k \in \{0, \dots, 9\}}$

(b) $$p(c \mid x^{(i)}, \theta, \pi) = \frac{P(x^{(i)} \mid c, \theta, \pi) \cdot P(c \mid \pi)}{P(x^{(i)})}$$

$$= \frac{P(x^{(i)} \mid c, \theta, \pi) \cdot P(c \mid \pi)}{\sum_c \left(P(x^{(i)} \mid c) \cdot P(c \mid \pi)\right)}$$

$$= \frac{\prod_j \left[\theta_{jc}^{x_j^{(i)}} (1-\theta_{jc})^{(1-x_j^{(i)})}\right] \cdot \pi_c}{\sum_c \left[\prod_j \theta_{jc}^{x_j^{(i)}} (1-\theta_{jc})^{(1-x_j^{(i)})} \cdot \pi_c\right]}$$

$$\log p(c \mid x^{(i)}, \theta, \pi) = \sum_j \left[x_j^{(i)} \log \theta_{jc} + (1-x_j^{(i)}) \log (1-\theta_{jc})\right] + \log \pi_c$$
$$- \log \sum_c \left[\prod_j \theta_{jc}^{x_j^{(i)}} (1-\theta_{jc})^{(1-x_j^{(i)})} \cdot \pi_c\right]$$

(c) The average log-likelihood per data point cannot be computed, since the probability can be so small such that it is rounded to 0. log 0 is undefined.

(d) Graph in jupyter notebook

(e) $P(\theta) = Beta(3,3)$

$$P(\theta|X) = \frac{P(X|\theta)\cdot P(\theta)}{P(X)} = \frac{P(X|C,\theta)\cdot P(\theta)}{P(X|C)}$$

$$Beta(\theta,3,3) = \underbrace{\frac{\Gamma(3+3)}{\Gamma(3)\Gamma(3)}}_{=D}\theta^2(1-\theta)^2$$

$$\frac{\partial \log P(\theta)}{\partial \theta_{ab}} = \frac{\partial \sum_j\sum_c \log \theta_{jc}^2}{\partial \theta_{ab}} + \frac{\partial \sum_j\sum_c \log(1-\theta_{jc})^2}{\partial \theta_{ab}} + \frac{\partial \sum_j\sum_c D}{\partial \theta_{ab}}=0$$

$$= \frac{2}{\theta_{ab}} + \frac{-2}{1-\theta_{ab}}$$

$$= \frac{2}{\theta_{ab}} - \frac{2}{1-\theta_{ab}}$$

Therefore $\frac{\partial \log P(\theta)}{\partial \theta} = M_2 \in \mathbb{R}^{784\times 10}$ and $M_{2jc} = -1 \quad \forall j,c$

$$\frac{\partial l(\theta|X)}{\partial \theta} = \frac{\partial \log P(\theta|X)}{\partial \theta} = \frac{\partial \log\left[\prod_i P(X^{(i)}|\theta)\cdot P(\theta)/P(X|C)\right]}{\partial \theta}$$

$$= \frac{\partial \sum_i l(X^{(i)}|\theta)}{\partial \theta} + \frac{\partial \log P(\theta)}{\partial \theta}$$

$$= M + \frac{\partial \log P(\theta)}{\partial \theta} = M + M_2$$

According (a) $M \in \mathbb{R}^{784\times 10}$ where $M_{jc} = \frac{\sum_{i=1}^{N} \mathbb{I}[X_j^{(i)}=1 \,\&\, c^{(i)}=c]}{\theta_{jc}} - \frac{\sum_{i=1}^{N} \mathbb{I}[X_j^{(i)}=0 \,\&\, c^{(i)}=c]}{1-\theta_{jc}}$

Therefore $\frac{\partial l(\theta|X)}{\partial \theta} = M + M_2 = M_3 \in \mathbb{R}^{784\times 10}$

$$M_{3jc} = 0 \iff M_{jc} + M_{2jc} = 0$$

$$\frac{\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=1 \,\&\, C^{(i)}=c]}{\theta_{jc}} - \frac{\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, C^{(i)}=c]}{1-\theta_{jc}} + \frac{2}{\theta_{jc}} - \frac{2}{1-\theta_{jc}} = 0$$

$$\frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=1 \,\&\, C^{(i)}=c]}{\theta_{jc}} = \frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, C^{(i)}=c]}{1-\theta_{jc}}$$

$$\frac{1}{\theta_{jc}} - 1 = \frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]}$$

$$\frac{1}{\theta_{jc}} = \frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]} + 1$$

$$\theta_{jc} = \frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}{4+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c] + \sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=1 \,\&\, c^{(i)}=c]}$$

$$\theta_{jc} = \frac{2+\sum_{i=1}^{N} \mathbb{I}[x_j^{(i)}=0 \,\&\, c^{(i)}=c]}{4+\sum_{i=1}^{N} \mathbb{I}[c^{(i)}=c]}$$

$$\theta_{jc} = \frac{2+\#\text{of sample where } j\text{th bit is 1 and class is } c}{4+\#\text{of sample where class is } c}$$

Question 2

(a) True (b) False

$$P(x_1, x_2) = P(x_1) \cdot P(x_2)$$

$$P(x_1, x_2 | c) = P(x_1|c) \cdot P(x_2|c)$$

$$P(x_1) = \sum_c P(x_1|c), \quad P(x_2) = \sum_c P(x_2|c)$$

$$P(x_1, x_2) = \sum_c P(x_1, x_2 | c)$$

$P(x_1) \cdot P(x_2) = \sum_c P(x_1|c) \cdot \sum_c P(x_2|c) \geq \sum_c P(x_1, x_2|c)$ Thus, (b) is false.