2.1 Suppose we have N objects $x^{(i)}$ with labels $c^{(i)}$,

The likelihood function can be represented as

$$L(\theta,\pi) = \prod_{i=1}^{n} P(x^{(i)}, c^{(i)} \mid \theta, \pi)$$

Then the log-likelihood function is

$$\ell(\theta,\pi) = \log(L) = \sum_{i=1}^{N} \log P(x^{(i)}, c^{(i)} \mid \theta,\pi)$$

$$= \sum_{i=1}^{N} \left[\log\left(P(c^{(i)} \mid \pi) \prod_{j=1}^{784} P(x_j^{(i)} \mid c^{(i)}, \theta)\right)\right]$$

Since each class label $c^{(i)}$ is a different item of clothing, and we have 10 different types,

$$P(c^{(i)} = c \mid \pi) = \prod_{c=0}^{9} \pi_c^{\mathbb{1}\{c^{(i)} = c\}}$$

Then, deriving the log-likelihood function,

$$\ell(\theta,\pi) = \sum_{i=1}^{N} \left[\log P(c^{(i)} \mid \pi) + \log\left(\prod_{j=1}^{784} P(x_j^{(i)} \mid c^{(i)}, \theta)\right)\right]$$

$$= \sum_{i=1}^{N} \left[\mathbb{1}\{c^{(i)}=c\} \sum_{c=0}^{9} \log \pi_c + \sum_{j=1}^{784} [x_j^{(i)} \log \theta_{jc} + (1-x_j^{(i)}) \log(1-\theta_{jc})]\right]$$

We know the MLE of π and θ are separable, so we can compute MLE of π_c and θ_{jc} independently.

MLE of θ_{jc}: before taking the derivative respect to θ_{jc}, we need to understand that the value of θ is determined by $j \in [1, 784]$ (pixels, integer) and $c \in [0,9]$ (class, integer only). to make the expression of θ_{jc} more rigorious, I'll use a new variable k to represent the pixel.

$$\frac{\partial}{\partial \theta_{jc}} \ell = \sum_{i=1}^{N} \sum_{k=1}^{784} \frac{\partial}{\partial \theta_{jc}} [x_k^{(i)} \log \theta_{kc^{(i)}} + (1-x_k^{(i)}) \log(1-\theta_{kc^{(i)}})]$$

when $k \neq j \Rightarrow$ this pixel doesn't contribute to parameter θ.

when $k = j$, the MLE is valid with $\mathbb{1}\{c^{(i)} = c\}$

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

Therefore, we solve above get

$$\boxed{\hat{\theta}_{jc} = \frac{\sum_{i=1}^{N} [x_j^{(i)} \mathbb{1}\{c^{(i)}=c\}]}{\sum_{i=1}^{N} \mathbb{1}\{c^{(i)} = c\}}}$$

MLE of π:

we still a new variable k, k has range $0 \sim 9$, represents class.

$$\frac{\partial}{\partial \pi_c} \ell = \sum_{i=1}^{N} \sum_{k=0}^{9} \frac{\partial}{\partial \pi_c} [\mathbb{1}(c^{(i)} = k) \log \pi_k]$$

$$= \sum_{i=1}^{N} \sum_{k=0}^{9} \frac{\partial}{\partial \pi_c} [\mathbb{1}\{c^{(i)} = k\} \log \pi_k + 1 - \sum_{k=0}^{9} \pi_k] \quad \text{since } \sum_{\text{all classes}} \pi_k = 1$$

when $k \neq c$, the derivative will be 0 because we take derivative respect to 0

$k = c$, $\frac{\partial}{\partial \pi_k} \pi_k = 1$

$$\text{so } \frac{\partial}{\partial \pi_c} \ell = \sum_{i=1}^{N} \left[\frac{\mathbb{1}\{c^{(i)}=c\}}{\pi_c} - 1\right] = 0 \Rightarrow \boxed{\hat{\pi}_c = \frac{\sum_{i=1}^{N} \mathbb{1}\{c^{(i)}=c\}}{N}}$$