CSC411 A6

Zhihong Wang 1002095207

2018-11

Part 1:

1. (similar to A4.Q2.d and A5.Q2.b)

   Based on the MAP formulation and distributions, we can set two L functions:

   Let $L(\pi) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_k^{(i)} log\pi_k - \lambda(\sum_{k=1}^{K} \pi_k - 1) + \sum_{k=1}^{K}(a_k - 1)log\pi_k$

   $L(\theta)$
   $= \sum_{i=1}^{N} r_k^{(i)}(x_j log\theta_{k,j} + (1-x_j)log(1-\theta_{k,j})) + (a-1)log\theta_{k,j} + (b-1)log(1-\theta_{k,j})$

   then $\frac{\partial L}{\partial \pi_k} = \frac{\sum_{i=1}^{N} r_k^{(i)}}{\pi_k} - \lambda + \frac{a_k - 1}{\pi_k}$

   $\frac{\partial L}{\partial \theta} = \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^i}{\theta_{k,j}} - \frac{\sum_{i=1}^{N} r_k^{(i)}(1-x_j^i)}{1-\theta_{k,j}} + \frac{a-1}{\theta_{k,j}} - \frac{b-1}{1-\theta_{k,j}}$

   First, let $\frac{\partial L}{\partial \pi_k} = 0$

   then $\lambda = \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\pi_k}$

   then $\pi_k = \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\lambda}$

   then $\sum_{k'=1}^{K} \pi_{k'} = \sum_{k'=1}^{K} \frac{a_{k'} - 1 + \sum_{i=1}^{N} r_{k'}^{(i)}}{\lambda}$

   since $\sum \pi_k = 1$

   then $\sum_{k'=1}^{K}(a_{k'} - 1 + \sum_{i=1}^{N} r_{k'}^{(i)}) = \lambda = \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\pi_k}$

   then $\pi_k = \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\sum_{k'=1}^{K}(a_{k'} - 1 + \sum_{i=1}^{N} r_{k'}^{(i)})}$

So $\pi_k \leftarrow \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\sum_{k'=1}^{K}(a_{k'} - 1 + \sum_{i=1}^{N} r_{k'}^{(i)})}$

Second, let $\frac{\partial L}{\partial \theta} = 0$

then $\frac{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1}{\theta_{k,j}} = \frac{\sum_{i=1}^{N} r_k^{(i)}(1 - x_j^{(i)}) + b - 1}{1 - \theta_{k,j}} = \frac{\sum_{i=1}^{N} r_k^{(i)} - \sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + b - 1}{1 - \theta_{k,j}}$

then $(1 - \theta_{k,j})(\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1) = \theta_{k,j}(\sum_{i=1}^{N} r_k^{(i)} - \sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + b - 1)$

then $\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1 = \theta_{k,j}(\sum_{i=1}^{N} r_k^{(i)} + b + a - 2)$

then $\theta_{k,j} = \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1}{\sum_{i=1}^{N} r_k^{(i)} + a + b - 2}$

So $\theta_{k,j} \leftarrow \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1}{\sum_{i=1}^{N} r_k^{(i)} + a + b - 2}$

Therefore,

$\pi_k \leftarrow \frac{a_k - 1 + \sum_{i=1}^{N} r_k^{(i)}}{\sum_{k'=1}^{K}(a_{k'} - 1 + \sum_{i=1}^{N} r_{k'}^{(i)})}$

$\theta_{k,j} \leftarrow \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1}{\sum_{i=1}^{N} r_k^{(i)} + a + b - 2}$

2.

```
pi[0] 0.085
pi[1] 0.13
theta[0, 239] 0.642710622711
theta[3, 298] 0.465736124958
```

Part 2:

1. (similar to A4.Q2.a)

$$Pr(z^i = k|x^i)$$

$$= \frac{p(z^i=k)p(m^i x^i|z^i=k)}{\sum_{k'=1}^{K} p(z^i=k')p(m^i x^i|z^i=k')}$$

$$= \frac{\pi_k \prod_{j=1}^{D} p(m_j^i x_j^i|z^i=k)}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j'=1}^{D} p(m_{j'}^i x_{j'}^i|z^i=k')}$$

$$= \frac{\pi_k \prod_{j=1}^{D} \theta_{k,j}^{m_j^i x_j^i} (1-\theta_{k,j})^{m_j^i (1-x_j^i)}}{\sum_{k'=1}^{K} \pi_{k'} \prod_{j'=1}^{D} \theta_{k',j'}^{m_{j'}^i x_{j'}^i} (1-\theta_{k',j'})^{m_{j'}^i (1-x_{j'}^i)}}$$

3.

```
R[0, 2] 0.174889514921
R[1, 0] 0.688537676109
P[0, 183] 0.651615199813
P[2, 628] 0.474080172491
```

Part 3:

1.
Assume $a = b = 1$, we can get $a - 1 = 0, a + b - 2 = 0$
then $\theta_{k,j} \leftarrow \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^i}{\sum_{i=1}^{N} r_k^{(i)}}$
If a pixel (i.e. x) is always 0 in training set, then the θ is also 0.
However, the θ represents the probability, which means it will assign 0 probability to the image in the test set.

2.
Because part 2 model has more classes (100 classes) than part 1 (10 classes). Since part 2 can predict with more criteria, the result of part 2 is more accurate than part 1.

3.
No. Because 1 has a special and unique upper half compared with other numbers. However, 8 shares similar top half with 9. Also, the heads of 0, 2 and 3 are similar to 8 under some conditions. Therefore, the images of 1's are assigned higher log probability than 8's.