Vele Tosevski

1002174657

CSC 411

March 27, 2019

HW6 Write-Up

Part 1

1. $$\sum_{i=1}^{N}\sum_{k=1}^{K} r_k^{(i)}\left[\log \Pr(z^{(i)}=k) + \log p(x^{(i)}|z^{(i)}=k)\right] + \log p(\pi) + \log p(\theta) \quad (1)$$

$\Pr(z^{(i)}=k) = \pi_k$ (2) since from multinomial dist.

$p(x^{(i)}|z^{(i)}=k) = \prod_{j=1}^{D} \theta_{kij}^{x_j^{(i)}}(1-\theta_{kij})^{1-x_j^{(i)}}$ (3) according to handout.

$p(\pi) = \prod_{j=1}^{K} \pi_j^{\alpha-1}$ (4) according to handout

$p(\theta) = \left[\theta_{kij}^{a-1}(1-\theta_{kij})^{b-1}\right]$ (5) according to handout. matrix

(2), (3), (4), (5) $\rightarrow$ (1)

$$\sum_{i=1}^{N}\sum_{k=1}^{K} r_k^{(i)}\left[\log \pi_k + \log \prod_{j=1}^{D} \theta_{kij}^{x_j^{(i)}}(1-\theta_{kij})^{1-x_j^{(i)}}\right] + \log \prod_{j=1}^{K} \pi_j^{\alpha-1} + \log\left[\theta_{kij}^{a-1}(1-\theta_{kij})^{b-1}\right]$$

$$\frac{\partial}{\partial \pi_k}(1) = \sum_{i=1}^{N}\left(r_k^{(i)}\frac{1}{\pi_k}\right) + (\alpha-1)\frac{1}{\pi_k} = 0 \quad \text{Let's use Lagrange to solve.}$$

$$\nabla f(a_k) = \frac{1}{\pi_k}\left[\left(\sum_{i=1}^{N} r_k^{(i)}\right) + \alpha - 1\right]$$

$$g(a_k) = 1 - \sum_{k'=1}^{K} \pi_{k'} \text{ since } \sum_{k'=1}^{K} \pi_{k'} = 1$$

$$\nabla g(\pi_k) = -1$$

$$\nabla f = \lambda \nabla g$$

$$\frac{\left(\sum_{i=1}^{N} r_k^{(i)}\right) + \alpha - 1}{\pi_k} = -\lambda$$

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

$$\therefore\ 1 - \sum_{k'=1}^{K} \pi_{k'} = 1 - \sum_{k'=1}^{K}\left[\frac{1 - \alpha - \sum_{i=1}^{N} r_{k'}^{(i)}}{\lambda}\right] = 0$$

$$\mapsto \lambda = \sum_{k'=1}^{K}\left[1 - \alpha - \sum_{i=1}^{N} r_{k'}^{(i)}\right] = K(1-\alpha) - \sum_{k'=1}^{K}\sum_{i=1}^{N} r_{k'}^{(i)}$$

$$\therefore\ \boxed{\pi_k = \frac{1 - \alpha - \sum_{i=1}^{N} r_k^{(i)}}{K(1-\alpha) - \sum_{k'=1}^{K}\sum_{i=1}^{N} r_{k'}^{(i)}}}$$

$$\frac{\partial}{\partial \theta_{k,j}}\left[① = \overbrace{\sum_{i=1}^{N}\sum_{k=1}^{K} r_k^{(i)}\left[\log \pi_k + \sum_{j=1}^{D}\left(x_j^{(i)} \log \theta_{k,j} + (1 - x_j^{(i)}) \log(1 - \theta_{k,j})\right)\right]}^{⑥} + \overbrace{\log \prod_{j=1}^{K} \pi_j^{\alpha - 1}}^{⑦} \right]$$

$$⑧ \left\{ \begin{array}{l} + (a-1)\log \theta_{k,j} \\ + (b-1)\log(1-\theta_{k,j}) \end{array} \right]$$

$$\frac{\partial}{\partial \theta_{k,j}} ⑥ = \sum_{i=1}^{N} r_k^{(i)}\left[\frac{x_j^{(i)}}{\theta_{k,j}} + (-)\frac{1 - x_j^{(i)}}{1 - \theta_{k,j}}\right] = \sum_{i=1}^{N} r_k^{(i)}\left[\frac{x_j^{(i)}}{\theta_{k,j}} + \frac{x_j^{(i)} - 1}{1 - \theta_{k,j}}\right]$$

$$\frac{\partial}{\partial \theta_{k,j}} ⑦ = 0$$

$$\frac{\partial}{\partial \theta_{k,j}} (8) = \frac{a-1}{\theta_{k,j}} + \frac{1-b}{1-\theta_{k,j}}$$

$$\frac{\partial}{\partial \theta_{k,j}} (1) = \frac{\partial}{\partial \theta_{k,j}} (6) + \frac{\partial}{\partial \theta_{k,j}} (7) + \frac{\partial}{\partial \theta_{k,j}} (8)$$

$$= \sum_{i=1}^{N} r_k^{(i)} \left[\frac{x_j^{(i)}}{\theta_{k,j}} + \frac{x_j^{(i)} - 1}{1-\theta_{k,j}} \right] + \frac{a-1}{\theta_{k,j}} + \frac{1-b}{1-\theta_{k,j}} = 0$$

$$= \frac{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)}}{\theta_{k,j}} + \frac{\sum_{i=1}^{N} r_k^{(i)} (x_j^{(i)} - 1)}{1-\theta_{k,j}} + \frac{a-1}{\theta_{k,j}} + \frac{1-b}{1-\theta_{k,j}} = 0$$

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

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

$$\theta_{k,j} \left[\sum_{i=1}^{N} r_k^{(i)} - \cancel{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)}} - 1 + b + \cancel{\sum_{i=1}^{N} r_k^{(i)} x_j^{(i)}} + a - 1 \right] = \sum_{i=1}^{N} r_k^{(i)} x_j^{(i)} + a - 1$$

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

$$\therefore \quad \boxed{\theta_{k,j} = \frac{\left[\sum_{i=1}^{N} \left(r_k^{(i)} x_j^{(i)} \right) \right] + a - 1}{\left[\sum_{i=1}^{N} \left(r_k^{(i)} \right) \right] + a + b - 2}}$$

2.

Code in mixture.py

Part 1 output →

pi[0] 0.085

pi[1] 0.13

theta[0, 239] 0.6427106227106232

theta[3, 298] 0.46573612495845823

The theta update seems OK.

The pi update seems OK.

Part 2

1. $p(z^{(i)} = k \mid x_{obs}^{(i)})$

$= p(z^{(i)} = k \mid X^{(i)})$ where $X = X_{obs}$ for simplicity

$= p(z^{(i)} = k, x^{(i)})$

$$= \frac{p(x^{(i)} \mid z^{(i)} = k)\, p(z^{(i)} = k)}{p(x^{(i)})}$$

$$= \frac{\pi_k \prod_{j=1}^{D} \theta_{k,j}^{x_j^{(i)}} (1-\theta_{k,j})^{1-x_j^{(i)}}}{\sum_{k'=1}^{K} \underbrace{p(z^{(i)} = k', x^{(i)})}}$$

$= p(x^{(i)} \mid z^{(i)} = k')\, p(z^{(i)} = k')$

$= \pi_{k'} \prod_{j=1}^{D} \theta_{k',j}^{x_j^{(i)}} (1-\theta_{k',j})^{1-x_j^{(i)}}$

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

2. In mixture.py

3. In mixture.py

Part 2 output →

R[0, 2] 0.17488951492117286

R[1, 0] 0.6885376761092291

P[0, 183] 0.6516151998131036

P[2, 628] 0.4740801724913303

The E-step seems OK.

Part 3

1. Images that have not appeared in the training set often will be given a probability of zero. That is due to the nature of the normal distribution. If a certain pixel is always 0 in the training set then if it is 1 in the test set, due to the normal distribution, it will be assigned a probability of 0 and the M-step will not match it to a correct cluster.

2. The part 2 model has partial observations so intuitively, it should perform better at every step.

3. If you sample from the distribution, you will get 1 and 8 and equal amount of times since the distribution should be equally likely for both of them. The model is only saying that it is easier to pick a 1 than an 8, not that it is more likely, so the answer is no.