

Homework3

刘喆骐 2020013163 探微化01

1 and 2

1. 1) $P(e, -s, -m, -b) = P(e) \cdot P(-m) \cdot P(-s|-m, -e) \cdot P(-b|-m)$

$P(-e) = 0.6 \quad P(-m) = 0.9 \quad P(-b|-m) = 0.9$

$P(-s|-m, -e) = 0.9 \Rightarrow P(e, -s, -m, -b) = 0.9^3 \times 0.6 = 0.4374$

2) $P(+b) = P(+b, +m) + P(+b, -m)$

$P(+b, +m) = P(+b|+m) \cdot P(+m) = 1 \times 0.1 = 0.1$

$P(+b, -m) = P(+b|-m) \cdot P(-m) = 0.1 \times 0.9 = 0.09$

$P(+b) = 0.1 + 0.09 = 0.19$

3) $P(+m|+b) = \frac{P(+m, +b)}{P(+b)} = \frac{0.1}{0.19} = \frac{10}{19}$

(4) -

2. 1) $T_2(b, e, f, g) = \sum_c T_1(b) P(c|b) P(d|c) P(e|c, d) P(f|c, d, g) \cdot P(g|c, f)$
条件图: $T_2(b, e, f, g)$

2) $T_3(b, d, f, g) = \sum_e T_2(b, d, f, g)$
条件图: $T_3(b, d, f, g)$

3) $T_4 = \sum_{b, d} T_3(b, d, f, g)$
条件图: $T_4(b, d, f)$

4) $P(b, d|f) = P(b, d, f) / P(f)$

$P(b, d, f) = \frac{T_4(b, d, f)}{\sum_b \sum_d T_4(b, d, f) = f}$

5) $A \rightarrow B$ 涉及 A, G, E, C.

$P(+m|+s, +b, +e) = P(+m|+b) \cdot P(+m|+s, +e)$

$P(+m|+s, +e) = \frac{P(+s|+e, +m) P(+m)}{P(+s|+e)} \quad P(+m|+b) = \frac{P(+b|+m) P(+m)}{P(+b)}$

$P(+s|+e) = P(+s|+e, +m) \cdot P(+m) + P(+s|+e, -m) \cdot P(-m) = \frac{1 \times 0.1}{1 \times 0.1 + 0.8 \times 0.9} = \frac{10}{82}$

$= 1 \times 0.1 + 0.8 \times 0.9 = 0.82$

$P(+m|+s, +b, +e) = 0.82 \times \frac{10}{19} = \frac{8.2}{19}$

5) S 未涉及, E, M 独立 $P(+e|+m) = P(+e) = 0.4$



$$3. \log p(D_n; w, r) = \sum_{i=1}^N \log \frac{1}{\sqrt{2\pi}} - \log \sigma - \frac{1}{2\sigma^2} (y_i - w^T x_i)^2.$$

$$\hat{f}(r, w) = \log p(D_n; w, r) \quad \hat{w} = \arg \min_w \sum_{i=1}^N (y_i - w^T x_i)^2.$$

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial \sigma} = 0 \Rightarrow -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^N (y_i - w^T x_i)^2 = 0 \\ \frac{\partial f}{\partial w_j} = 0 \Rightarrow \sum_{i=1}^N (y_i - w^T x_i) x_{ij} = 0 \end{array} \right.$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^N (y_i - w^T x_i)^2}{N}.$$

$$\hat{w} = (X^T X)^{-1} X^T y, \quad X = [\vec{x}_1, \dots, \vec{x}_N]$$

$$\hat{w}^T = \frac{\sum_{i=1}^N y_i x_i}{\sum_{i=1}^N x_i^2}, \quad \hat{w} = \frac{\sum_{i=1}^N y_i x_i^T}{\sum_{i=1}^N |x_i|^2}, \quad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$w^T X = \hat{w}^T [\vec{x}_1, \dots, \vec{x}_N] = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \vec{y}.$$

$$\Rightarrow \hat{w}^T = \vec{y} X^{-1}.$$

$$\hat{w} = [\vec{y} X^{-1}]^T.$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^N (y_i - [(X^T X)^{-1} X^T y]^T x_i)^2}{N}.$$

$$4. 1) \alpha_{(2,2),(1,3)} = \min\left(1, \frac{p(1,3)Q_{(1,3)}(2,2)}{p(2,2)Q_{(2,2)}(1,3)}\right)$$

均匀分布 $\rightarrow Q = \text{const.}$ $p(x_1, x_2) \propto x_1 + \ln(x_1 x_2 + 2x_1 + 3x_2)$

$$\alpha_{(2,2),(1,3)} = \min\left(1, \frac{p(1,3)}{p(2,2)}\right)$$

$$\frac{p(1,3)}{p(2,2)} = \frac{1 + \ln(1 \times 3 + 2 \times 1 + 3 \times 3)}{2 + \ln(2 \times 2 + 2 \times 2 + 3 \times 2)} = \frac{1 + \ln 14}{2 + \ln 14}.$$

$$\Rightarrow \alpha_{(2,2),(1,3)} = 1.$$

$$\alpha_{(2,2),(3,4)} = \frac{p(3,4)}{p(2,2)}$$

$$\frac{p(3,4)}{p(2,2)} = \frac{3 + \ln(3 \times 4 + 2 \times 3 + 3 \times 4)}{2 + \ln(2 \times 2 + 2 \times 2 + 3 \times 2)} = \frac{3 + \ln 30}{2 + \ln 14}.$$

$$\Rightarrow \alpha_{(2,2),(3,4)} = \frac{3 + \ln 30}{2 + \ln 14}.$$

$$2) \cancel{q(x_1)} \cancel{p(x_1|x_2)}, \cancel{p(x_1|x_2)} = \frac{p(x_1, x_2)}{\int_{x_2} p(x_1, x_2) dx_2}$$

$q(x_1)$ 的形式取决于 $p(x_1|x_2)$, 在 $p(x_1|x_2)$ 中采样.

$$\text{而 } p(x_1, x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{p(x_1, x_2)}{\int_{x_1} p(x_1, x_2) dx_1}$$

取 $x_1 = x_2$ 时, 采样 x_2

$q(x_2)$ 取决于 $p(x_2|x_1)$

$$p(x_2, x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{p(x_1, x_2)}{\int_{x_2} p(x_2, x_1) dx_2}$$

4) $J(\theta) = \sum_z q^*(z) \log \frac{P(x, z | \theta)}{q^*(z)}$ $q^*(z) = P(z | x, \theta^{old})$

* $\arg \max_{\theta} J(\theta) = \arg \max_{\theta} \sum_z q^*(z) \log P(x, z | \theta) - \arg \max_{\theta} \sum_z q^*(z) \log q^*(z) \propto \arg \max_{\theta} \sum_z P(z | x, \theta^{old}) \log P(x, z | \theta)$

$= \arg \max_{\theta} \sum_z \frac{P(z, x | \theta^{old})}{P(x | \theta^{old})} \log P(x, z | \theta) = \frac{1}{P(x | \theta^{old})} \arg \max_{\theta} \sum_z P(z, x | \theta^{old}) \log P(x, z | \theta)$

6) $P(x, z | \theta) = \pi_{z_1} \left[\prod_{t=1}^{T-1} b_{z_t}(x_t) a_{z_t, z_{t+1}} \right] b_{z_T}(x_T) = \pi_{z_1} \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^T b_{z_t}(x_t)$

3) $Q(\theta | \theta^{old}) = \sum_z P(x, z | \theta^{old}) \log \pi_{z_1} + \sum_z P(x, z | \theta^{old}) \sum_{t=1}^{T-1} (\log a_{z_t, z_{t+1}} + \sum_z P(x, z | \theta^{old}) \sum_{t=1}^T \log b_{z_t}(x_t))$

4) $\sum_z P(x, z | \theta^{old}) \log \pi_{z_1} = \sum_{i=1}^N \log \pi_{z_1=i} P(x, z_1=i | \theta^{old}) = \sum_{i=1}^N P(x, z_1=i | \theta^{old}) \log \pi_{z_1}$

$\mathcal{L} = \sum_{i=1}^N \log \pi_{z_1} P(x, z_1=i | \theta^{old}) + \lambda (\sum_{i=1}^N \pi_{z_1=i} - 1) \quad \frac{\partial \mathcal{L}}{\partial \pi_{z_1}} = \frac{P(x, z_1=i | \theta^{old})}{\pi_{z_1}} + \lambda = 0 \Rightarrow \pi_{z_1} = \frac{P(x, z_1=i | \theta^{old})}{\sum_{i=1}^N P(x, z_1=i | \theta^{old})}$

$\Rightarrow \lambda = - \sum_{i=1}^N P(x, z_1=i | \theta^{old}) = -P(x | \theta^{old}) \Rightarrow \pi_{z_1} = \frac{P(x, z_1=i | \theta^{old})}{P(x | \theta^{old})}$

5) $\mathcal{L}_1 = \sum_z P(x, z | \theta^{old}) \sum_{t=1}^{T-1} \log a_{z_t, z_{t+1}} + \lambda (\sum_j a_{ij} - 1) \quad \frac{\partial \mathcal{L}_1}{\partial a_{ij}} = \frac{\sum_{t=1}^{T-1} P(x, z_t=i, z_{t+1}=j | \theta^{old})}{a_{ij}} + \lambda = 0$

$\Rightarrow \sum_j a_{ij} = 1 \Rightarrow a_{ij} = \frac{\sum_{t=1}^{T-1} P(x, z_t=i, z_{t+1}=j | \theta^{old})}{\sum_{t=1}^{T-1} P(x, z_t=i | \theta^{old})}$

$\mathcal{L}_2 = \sum_{k=1}^M P(x, z_t=j | \theta^{old}) \sum_{t=1}^T \log b_{z_t}(x_t) + \lambda (\sum_k b_j(k) - 1)$

$\frac{\partial \mathcal{L}_2}{\partial b_j(k)} = \frac{P(x, z_t=j, x_t=k | \theta^{old})}{b_j(k)} + \lambda \Rightarrow b_j(k) = \frac{\sum_{t=1}^T P(x, z_t=j, x_t=k | \theta^{old})}{\sum_{t=1}^T P(x, z_t=j | \theta^{old})}$

6) $\gamma_t^{old}(i) = P(z_t=i | x, \theta^{old}) = \frac{P(z_t=i, x | \theta^{old})}{P(x | \theta^{old})} = \frac{P(z_t=i, x | \theta^{old})}{\sum_j P(z_t=j, x | \theta^{old})} = \frac{\alpha_t^{old}(i) \beta_t^{old}(i)}{\sum_j \alpha_t^{old}(j) \beta_t^{old}(j)}$

$\xi_t(i, j) = P(z_t=i, z_{t+1}=j | x, \theta) = \frac{P(z_t=i, z_{t+1}=j, x | \theta^{old})}{P(x | \theta^{old})} = \frac{P(z_t=i, z_{t+1}=j, x | \theta^{old})}{\sum_{i,j} P(z_t=i, z_{t+1}=j, x | \theta^{old})} = \frac{\alpha_t^{old}(i) \beta_t^{old}(j)}{\sum_{i,j} \alpha_t^{old}(i) \beta_t^{old}(j)}$

7) $\pi_i = \frac{P(x, z_1=i | \theta^{old})}{P(x | \theta^{old})} = \pi_i^{old}$

$a_{ij} = \frac{\sum_{t=1}^{T-1} P(x, z_t=i, z_{t+1}=j | \theta^{old})}{\sum_{t=1}^{T-1} P(x, z_t=i | \theta^{old})} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j) P(x | \theta^{old})}{\sum_{t=1}^{T-1} P(x, z_t=i | \theta^{old})} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t^{old}(i)}$

$b_j(k) = \frac{\sum_{t=1}^T P(x, z_t=j, x_t=k | \theta^{old})}{\sum_{t=1}^T P(x, z_t=j | \theta^{old})} = \frac{\sum_{t=1}^T \delta_t^{old}(j) 1(x_t=k)}{\sum_{t=1}^T P(x, z_t=j | \theta^{old})} = \frac{\sum_{t=1}^T \delta_t^{old}(j) 1(x_t=k)}{\sum_{t=1}^T \gamma_t^{old}(j)}$

伪代码如下:

```
def Baum-Welch(X, pi, A, B, pi_old, A_old, B_old, max_iter)
    model.init(pi_old, A_old, B_old)
    for i in range(max_iter):
        alpha = model.forward(X)
        beta = model.backward(X)
        gamma = model.compute_gamma(alpha, beta)
        xi = model.compute_xi(alpha, beta)
        pi = model.compute_p(gamma)
        A = model.compute_a(gamma, xi)
        B = model.compute_b(gamma)
        model.update(pi, A, B)
    return model
```

6

(a)

伪代码如下:

```
def VariationalEM(text,max_iter)
    docs, vocab = Preprocess(text)    lda_model = LDA()
    lda_model.fit(docs, vocab, max_iter)
    init_parameters()
    while (i < max_iter and delta < tolerance)
        for doc in docs:
            update_phi(doc)
            update_gamma(doc)
        update_log_beta()
        update_alpha()
        prev_bound = bound
        update_bound()
        delta = 1- bound / prev_bound
    print_results()
```

(b)

见代码文件

(c)

分别取K=5,10,20,结果如下:

K=5

0	孩子	发现	调查	警方	现场	民警	情况	医院
1	男子	医院	发现	孩子	发生	民警	介绍	警方
2	男子	公司	孩子	警方	法院	情况	发现	女士
3	民警	警方	孩子	派出所	告诉	嫌疑人	报警	两个
4	发现	孩子	学生	学校	老师	家长	警方	下午

K=10

0	孩子	男子	老人	学生	儿子	手机	发生	发现
1	发现	孩子	工作人员	现场	介绍	车辆	调查	情况
2	民警	警方	派出所	孩子	法院	情况	发现	万元
3	发现	孩子	警方	法院	男子	证据	时间	相关
4	公司	工作	孩子	万元	找到	银行	女儿	家人
5	视频	孩子	网友	告诉	学生	标题	发现	显示
6	民警	孩子	男子	李某	女子	警方	发现	学生
7	民警	男子	孩子	发生	编辑	警方	发现	人员
8	医院	孩子	发生	情况	调查	医生	发现	告诉
9	老人	警方	男子	医院	女士	民警	下午	发现

K=20

0	孩子	车辆	下午	医院	司机	民警	女儿	死亡
1	孩子	老师	家长	学生	公司	发现	照片	手机
2	男子	手机	警方	民警	嫌疑人	公司	发现	标题
3	学生	事件	发生	调查	发现	学校	孩子	相关
4	编辑	警方	民警	发现	孩子	男子	公司	法院
5	孩子	告诉	医生	医院	发生	工作	手术	发现
6	法院	万元	民警	医院	公司	发现	工作人员	警方
7	发现	村民	调查	民警	男子	昨日	警方	小区
8	发现	医院	民警	公司	报道	李桂英	调查	死亡
9	警方	孩子	医院	民警	儿子	发现	现场	调查
10	警方	父亲	男子	民警	发现	派出所	情况	现场
11	男子	工作	民警	公司	希望	老人	网友	警方
12	男子	民警	报警	现场	发现	派出所	医院	发生
13	两人	民警	公司	标题	发现	警方	医院	老人
14	民警	发现	警方	李某	电话	工作	调查	情况
15	男子	警方	报道	标题	民警	女子	发现	情况
16	孩子	学生	学校	女士	家长	发现	告诉	相关
17	发现	孩子	警方	民警	医院	男子	工作	公司
18	介绍	发现	孩子	老人	时间	民警	公司	相关
19	女士	儿子	警方	发现	学校	老师	孩子	发生

(d)

最好的K为10。当K过小时，分类的主题数小于真实的主题数，导致部分主题缺失，对应欠拟合。而K过大时，分类的主题数大于真实的主题数，使得相同主题反复出现以及无义词被组合成一个主题，对应过拟合。于是K过大和过小都不好。