张展玮 12110817

Question 1 (a)

使用EM算法：

1. **期望步骤（E-step）**：计算每个数据点属于每个高斯分量的期望概率（即后验概率），这些概率称为责任（responsibilities）。给定当前参数估计，对于数据集中的每个点 x_i，我们计算它来自第 k 个分量的概率：

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}$$

   其中 $\gamma(z_{ik})$ 表示第 i 个数据点来自第 k 个分量的责任。

2. **最大化步骤（M-step）**：
   - 更新每个分量的混合权重：

$$\pi_k^{new} = \frac{1}{N}\sum_{i=1}^{N}\gamma(z_{ik})$$

   - 更新每个分量的均值：

$$\mu_k^{new} = \frac{\sum_{i=1}^{N}\gamma(z_{ik})x_i}{\sum_{i=1}^{N}\gamma(z_{ik})}$$

   - 更新每个分量的协方差矩阵：

$$\Sigma_k^{new} = \frac{\sum_{i=1}^{N}\gamma(z_{ik})(x_i-\mu_k^{new})(x_i-\mu_k^{new})^T}{\sum_{i=1}^{N}\gamma(z_{ik})}$$

Question 1 (b)

1. **混合权重 π_k 的更新**，考虑狄利克雷先验：

$$\pi_k^{new} = \frac{\sum_{i=1}^{N}\gamma(z_{ik}) + N_{0k} - 1}{N + \sum_{k=1}^{K}(N_{0k}-1)}$$

   其中 N_{0k} 是狄利克雷分布的先验参数。

2. **均值 μ_k 的更新**，考虑正态先验：

$$\mu_k^{new} = \frac{\sum_{i=1}^{N}\gamma(z_{ik})x_i + \lambda\mu_{mk0}}{\sum_{i=1}^{N}\gamma(z_{ik}) + \lambda}$$

   其中 λ 是由先验协方差 Σ_{k0} 决定的缩放因子，μ_{mk0} 是先验均值。

3. **协方差矩阵 Σ_k 的更新**，考虑先验：

$$\Sigma_k^{new} = \frac{\sum_{i=1}^{N} \gamma(z_{ik})(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T + S_0}{\sum_{i=1}^{N} \gamma(z_{ik}) + \nu_0 + D + 2}$$

其中 S_0 是先验协方差矩阵，ν_0 是与自由度相关的先验参数，D 是数据维度。

Question 1 (c)

$$P(x_{N+1}|\theta_{MAP}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x_{N+1}|\mu_k, \Sigma_k)$$

Question 2 (a)

在隐马尔可夫模型（HMM）中，Baum-Welch 算法的 E 步骤和 M 步骤可以更详细地描述如下：

E步骤（Expectation）

E步骤的目的是计算隐状态的期望，即对每个时间点和每对状态，计算在当前参数下观测数据与隐状态序列相关联的概率。

1. **计算前向概率** $\alpha_t(i)$：这是在时间点 t 观测到 $x_1, \ldots, x_t$ 并且状态为 i 的联合概率。

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) A_{ij} \right] b_j(x_{t+1})$$

   其中，A_{ij} 是状态转移概率，$b_j(x_{t+1})$ 是在状态 j 观测到 x_{t+1} 的概率。初始前向概率 $\alpha_1(i)$ 是用初始状态分布 π_i 乘以在状态 i 观测到 x_1 的概率 $b_i(x_1)$。

2. **计算后向概率** $\beta_t(i)$：这是在时间点 t 状态为 i 并且在 t 之后观测到 $x_{t+1}, \ldots, x_N$ 的概率。后向概率通过下面的递归关系计算：

$$\beta_t(i) = \sum_{j=1}^{N} A_{ij} b_j(x_{t+1}) \beta_{t+1}(j)$$

   最后一个观测之后的后向概率 $\beta_N(i)$ 被初始化为 1。

3. **计算状态占用概率** $\gamma_t(i)$：这是在时间点 t 状态为 i 的概率。它可以通过前向和后向概率计算得到：

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$$

4. **计算状态转移概率** $\xi_t(i,j)$：这是在时间点 t 状态为 i 并且在时间点 $t+1$ 状态为 j 的概率。它可以由前向概率、后向概率和状态转移概率计算得到：

$$\xi_t(i,j) = \frac{\alpha_t(i) A_{ij} b_j(x_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^{N} \sum_{l=1}^{N} \alpha_t(k) A_{kl} b_l(x_{t+1}) \beta_{t+1}(l)}$$

M步骤（Maximization）

M步骤的目的是使用 E 步骤中计算的期望值来更新参数，最大化完全数据的对数似然。

1. **更新初始状态分布** π：

$$\pi_i = \gamma_1(i)$$

2. **更新状态转移矩阵** A：

$$A_{ij} = \frac{\sum_{t=1}^{N-1} \xi_t(i,j)}{\sum_{t=1}^{N-1} \gamma_t(i)}$$

3. **更新观测模型参数**：对于高斯观测模型，均值 μ_k 和协方差矩阵 Σ_k 的更新是基于 $\gamma_t(i)$ 的加权平均：

$$\mu_k = \frac{\sum_{t=1}^{N} \gamma_t(k) x_t}{\sum_{t=1}^{N} \gamma_t(k)}$$

$$\Sigma_k = \frac{\sum_{t=1}^{N} \gamma_t(k)(x_t - \mu_k)(x_t - \mu_k)^T}{\sum_{t=1}^{N} \gamma_t(k)}$$

Question 2 (b)

E步骤（Expectation）

E步骤保持不变，因为在此步骤中计算的前向概率（α）和后向概率（β）不涉及参数的先验信息。

M步骤（Maximization）

M步骤需要调整以包括先验信息。具体来说，先验信息将作为正则化项添加到参数更新公式中：

1. **初始状态分布** π **的更新**：

$$\pi_i = \frac{\sum_{t=1}^{N} \gamma_t(i) + N_{0i} - 1}{N + \sum_{j=1}^{K} (N_{0j} - 1)}$$

2. **转移矩阵** A **的更新**：

$$A_{ij} = \frac{\sum_{t=1}^{N-1} \xi_t(i,j) + M_{0ij} - 1}{\sum_{t=1}^{N-1} \gamma_t(i) + \sum_{j=1}^{K} (M_{0ij} - 1)}$$

3. **观测模型参数** μ_k **和** Σ_k **的更新**：

$$\mu_k = \frac{\sum_{t=1}^{N} \gamma_t(k) x_t + \Sigma_{k0}^{-1} \mu_{mk0}}{\sum_{t=1}^{N} \gamma_t(k) + \Sigma_{k0}^{-1}}$$

Question 2 (c)

$$P(x_{N+1} | \theta_{MAP}) = \sum_{z_{N+1}} P(x_{N+1} | z_{N+1}, \theta_{MAP}) P(z_{N+1} | D, \theta_{MAP})$$

Question 3 (a)

```
(0.11932500000000001,
 array([[0.1      , 0.039    , 0.0954   ],
        [0.45     , 0.3195   , 0.023925]]))
```

$$p(D|\pi, A, B) = 0.119325$$

Question 3 (b)

```
(array([[0.1      , 0.039    , 0.0954   ],
        [0.45     , 0.3195   , 0.023925]]),
 array([[0.174 , 0.52  , 1.    ],
        [0.2265, 0.31  , 1.    ]]),
 array([[0.14582024, 0.169956  , 0.79949717],
        [0.85417976, 0.830044  , 0.20050283]]),
 array([1, 1, 0]),
 0.057024750000000006)

array([[0.15688246, 0.01307354],
       [0.64261471, 0.18742929]])
```

$$P(Z_1 = bull) = 0.14582024$$
$$P(Z_2 = bull) = 0.169956$$
$$P(Z_3 = bull) = 0.79949717$$
$$P(Z_2 = \text{bull}, Z_3 = \text{bull}|D) = 0.15688246$$
$$P(Z_2 = \text{bull}, Z_3 = \text{bear}|D) = 0.01307354$$
$$P(Z_2 = \text{bear}, Z_3 = \text{bull}|D) = 0.64261471$$
$$P(Z_2 = \text{bear}, Z_3 = \text{bear}|D) = 0.18742929$$

Question 3 (c)

bear, bear, bull

Question 3 (d)

$$P(x_4 = rise) = 0.05702475$$