

# ESC 2024 Spring Session Week 2

## Mixture Models and EM

이상윤, 이현우

### 1. Introduction

Gaussian Mixture(GM) Model은 데이터를 여러 개의 가우시안 분포의 결합으로 모델링하는 방법입니다. EM(Expectation-Maximization) 알고리즘을 사용하면, GM 모델을 효과적으로 학습시킬 수 있습니다.

### 2-1. Gaussian Mixture Model

GM 모델은 다음과 같이 정의됩니다.

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k)$$

여기서  $\pi_k$ 는 k번째 가우시안 분포의 coefficient이며,  $\mu_k$ 와  $\Sigma_k$ 는 각각 k번째 가우시안 분포의 평균과 공분산을 나타냅니다.

### 2-2. EM Algorithm

EM 알고리즘은 GM 모델의 파라미터를 최적화하기 위해 사용됩니다. 이 알고리즘은 두 단계, 즉 Expectation 단계와 Maximization 단계를 반복 수행합니다. E 단계에서는 현재 모델 파라미터를 바탕으로 각 데이터 포인트가 각 가우시안 component에 속할 확률을 계산하고(soft assignment), M 단계에서는 이 확률을 사용하여 모델의 파라미터를 업데이트합니다.

### 2-3. Initialization

초기화 단계에서는 각 가우시안 component의 평균, 공분산, 그리고 각 component의 coefficient을 나타내는 파라미터들을 초기화합니다. 이 값들은 알고리즘의 결과에 영향을 줄 수 있기 때문에, 여러 번 초기화하여 다양한 결과를 얻는 것이 좋습니다.

### 2-4. Expectation Step

E 단계에서는 현재의 파라미터를 사용하여 각 데이터 포인트가 각 가우시안 component에 속할 확률을 계산합니다. 이 확률은 responsibility라고 하며, 다음과 같이 계산됩니다.

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

여기서  $\mathcal{N}(x | \mu, \Sigma)$ 는 가우시안 분포의 확률 밀도 함수를 나타냅니다.

## 2-5. Maximization Step

M 단계에서는 E 단계에서 계산한 책임도를 바탕으로 모델의 파라미터를 업데이트합니다. 새로운 파라미터 값은 다음과 같이 계산됩니다.

$$\mu_k^{(new)} = \frac{\sum_{i=1}^N r_{ik} x_i}{\sum_{i=1}^N r_{ik}} \quad (3)$$

$$\Sigma_k^{(new)} = \frac{\sum_{i=1}^N r_{ik} (x_i - \mu_k^{(new)})(x_i - \mu_k^{(new)})^T}{\sum_{i=1}^N r_{ik}} \quad (4)$$

$$\pi_k^{(new)} = \frac{\sum_{i=1}^N r_{ik}}{N} \quad (5)$$

## 2-6. Evaluating Log-likelihood

각 반복 후, 모델의 로그 가능도를 계산하여 알고리즘이 수렴했는지 평가합니다. 로그 가능도는 다음과 같이 계산됩니다.

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{i=1}^N \ln \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \right) \quad (6)$$

이 과정을 통해 데이터를 잘 설명하는 가우시안 혼합 모델의 파라미터를 찾을 수 있습니다

## 3-1. EM Algorithm을 사용하는 목적

- : 잠재 변수(latent variables)를 갖는 모델의 MLE 해를 구하기 위해서
- : 잠재 변수 Z에 대해 사전에 알려진 정보는 없으며, Z와 입력변수 X에 대한 조건부 분포 형태로 문제를 풀어야 합니다.

## 3-2. 잠재 변수가 주어진 모델의 MLE (Complete Data. vs Incomplete Data)

$$\ln p(\mathbf{X}|\theta) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\theta) \right\}$$

잠재변수가 주어진 경우, 아무리  $p(X, Z|\theta)$ 가 지수족 분포를 따른다고 해도, 로그함수 내에 합의 형태로 주어졌기 때문에 이에 대한 주변 확률 분포인  $p(X|\theta)$ 는 지수족 분포를 따르지 않을 수 있습니다. 따라서 복잡한 형태의 MLE 풀이를 갖게 되고, 이를 편리하게 해주는 것이 EM Algorithm입니다.

모든 관찰 데이터 X에 대해 잠재 변수 Z가 주어진 Complete Data {X, Z}의 경우에는 각각의 경우에 대해 로그 가능도 함수를 구하면 되므로 어렵지 않지만, 실제로 주어지게 되는 관찰 데이터에서는 일반적으로 잠재변수의 실제 값을 얻을 수 없습니다. 또한 가능한 모든 잠재변수의 값에 대한 분포를 고려해야 하기 때문에 로그 가능도 함수는 아래와 같고,

$$\ln p(X|\theta) = \ln \sum_Z p(X, Z|\theta)$$

로그 내에 합의 형태를 갖는 수식이 존재하므로 최대화 지점을 찾기 어렵습니다.

따라서 MLE를 풀기 위해서 다음 방식을 고려해볼 수 있습니다.

- : 제공된 샘플을 통해 잠재 변수를 추정하여 값이 제공되었다고 생각하여 계산
- : 이때, 잠재 변수의 값은 정확하게 알기 어렵기 때문에 기대값을 활용 ( $E[\ln p(X, Z|\theta)]$ )
- : 이때 이 기대값을 최대로 만드는 파라미터를 추론하는 과정이 EM Algorithm

#### · E-step

주어진 파라미터  $\theta^{old}$ 는 고정되어 있고, 이를 활용해 잠재변수  $Z$ 의 확률값  $p(Z|X, \theta^{old})$ 를 얻어야 한다.  
이제 다시  $p(Z|X, \theta^{old})$ 를 활용하여 complete-data log likelihood의 expectation을 서술하면 다음과 같다.

$$Q(\theta, \theta^{old}) = \sum_Z p(Z|X, \theta^{old}) \ln p(X, Z|\theta). \quad (9.30)$$

#### · M-step

예측된  $Z$ 의 기대값을 최대화하는 새로운 파라미터  $\theta^{new}$ 를 구한다.

$$\theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{old}). \quad (9.31)$$

· 이 과정을 반복하여, 결과값이 수렴할 경우 종료한다.

▶ EM Algorithm

### 3-3. 잠재 변수를 이용한 GMM의 재해석

: PPT p.24 ~ 27 참고

### 3-4. K-means 알고리즘 vs EM 알고리즘

K-means 알고리즘은 하나의 샘플이 오로지 하나의 클러스터에만 속하는 구조이고, 각 클러스터에 대한 분산도는 고려하지 않습니다. 이에 반해 EM Algorithm은 각각의 클러스터에 대해 샘플이 속할 확률 값을 표현한다는 점에서 차이가 있습니다.

### 3-5. 베르누이 혼합 분포 (with EM Algorithm)

: PPT p.30 ~ 33 참고

### 4-1. variational inference (변분 추론)

:  $q(Z)$ 를 도입하여 ELBO를 최대화함으로써 계산 불가능한 KL-divergence를 간접적으로 줄이는 방법

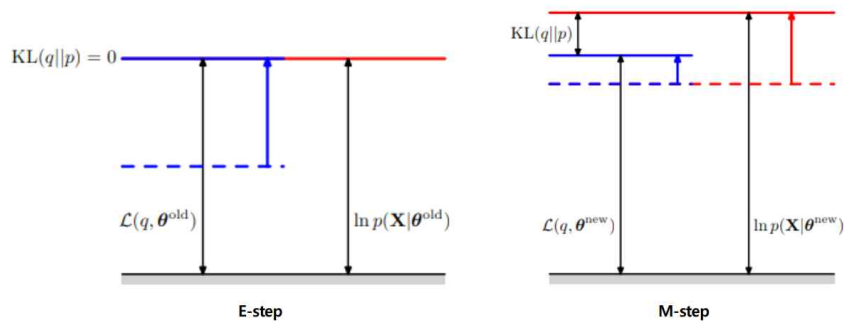
추론의 목적은 데이터의 likelihood를 계산하고, 잠재 변수  $Z$ 의 posterior distribution  $p(Z|X)$ 를 구하는 것이다. 이때 정확한 분포를 알지는 못하므로, 최적화 문제로 바꾸어 최대한 비슷한 값을 구하는 것

$$\begin{aligned}\log p(\mathbf{X}) &= \int q(\mathbf{Z}) \log \frac{p(\mathbf{X}, \mathbf{Z})}{q(\mathbf{Z})} d\mathbf{Z} - \int q(\mathbf{Z}) \log \frac{p(\mathbf{Z}|\mathbf{X})}{q(\mathbf{Z})} d\mathbf{Z} \\ &= \mathcal{L}(q) + \text{KL}(q(\mathbf{Z}) \| p(\mathbf{Z}|\mathbf{X}))\end{aligned}$$

- 1)  $\mathbf{Z}$ 는 잠재 변수,  $q$ 는 다를 수 있는  $\mathbf{Z}$ 를 확률변수로 갖는 임의의 확률분포
- 2)  $\mathcal{L}$ 은 ELBO(Evidence Lower Bound) 또는 Variational Free Energy라고 한다.
- 3) 위의 항 KL은 분포 간의 유사도의 척도로 사용되는 KL-divergence ( $q = p$ 일 때  $KL = 0$ )

## 4-2. The EM Algorithm in General

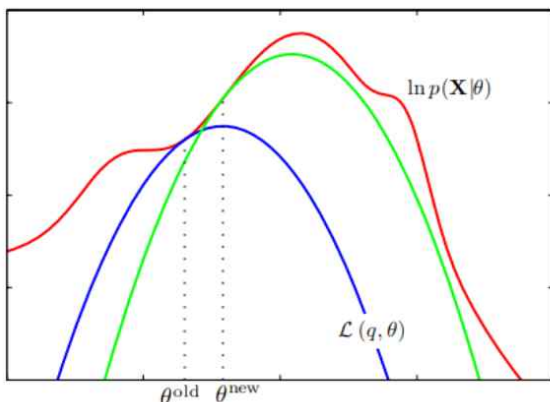
- : EM Algorithm은 잠재 변수를 갖는 확률 모델에서 MLE를 구하는 일반적인 기법
- : 또한 변분 추론(variational inference)의 한 방식이다



- 1) E-step  
:  $L(q, \theta)$ 에서  $\theta$ 를 고정( $\theta^{old}$ ),  $L(q, \theta^{old})$ 의 값을 최대화하는  $q$  함수를 선택하는 단계
- 2) M-step  
:  $q(\mathbf{Z})$ 를 고정시킨 상태에서  $L(q^{fixed}, \theta)$ 의 값을 최대화하는 새로운 파라미터  $\theta^{new}$ 를 추론하는 단계

## 4-3. Summary : EM Algorithm

- EM 알고리즘은 반복적인 과정을 통해 MLE를 구하는 알고리즘
- :  $q$  함수를 도입하여 일반화된 EM 알고리즘을 도출할 수 있다.



### E-step

- 1) 임의로 고정한  $\theta^{old}$ 로부터 EM 알고리즘 시작
- 2) 해당 지점에서 로그 가능도 함수와 최대한 근사한  $\mathcal{L}$  함수를 만든다.

### M step

- 3) 얻어진  $\mathcal{L}$  함수를 최대화하는 새로운 파라미터를 선정한다.

### Then..

- 4) 수렴 조건을 만족할 때까지 반복