



Foundation of Machine Learning 14주차

정재현, 우지수 / 2023.06.28



Computational Data Science LAB

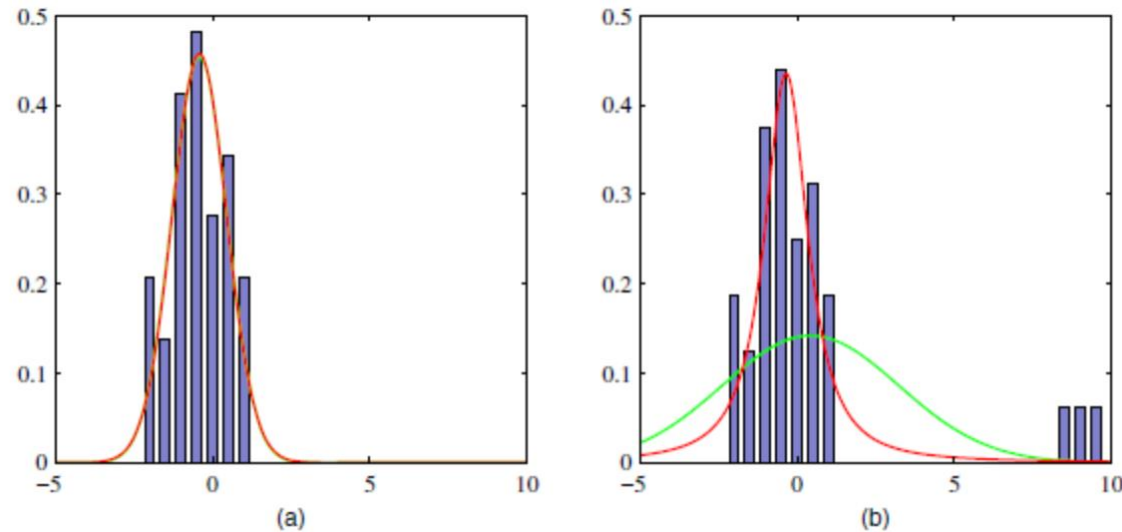
CONTENTS

1. Introduction
2. Gaussian Mixture Model
3. Learn the Gaussian Mixture Model
4. EM Algorithm for Latent Variable Models
5. EM Algorithm for Gaussian Mixture Model
6. Gaussian Mixture Model for Clustering

01 | Introduction

Gaussian Mixture Model Background

- Why we use Gaussian Mixture Model?

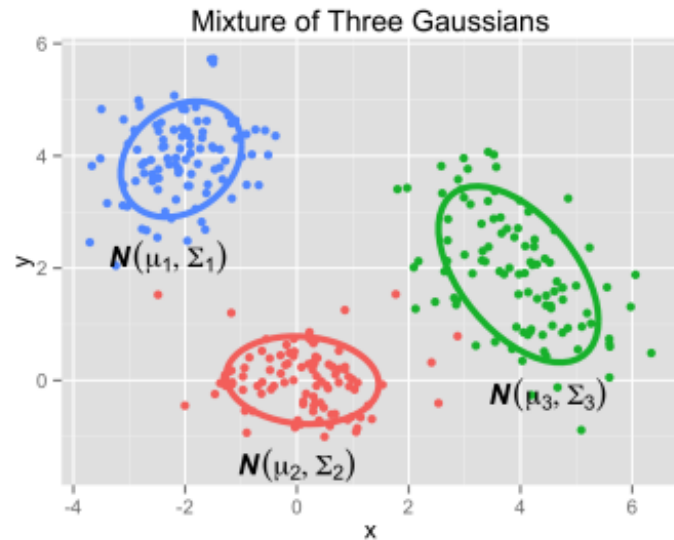


- ✓ (a)와 달리 (b)는 하나의 Gaussian distribution으로 표현하기 어렵다는 문제 존재
- ✓ 또한, 모든 데이터가 종 모양 분포를 띄진 어렵기 때문에 Gaussian distribution은 실제 데이터셋을 모델링하는 데에 한계 존재
- ✓ 이에 Gaussian distribution을 선형 결합하여 만든 분포인 Gaussian Mixture Model 등장

02 | Gaussian Mixture Model

Theory

- Gaussian Mixture Model (GMM)



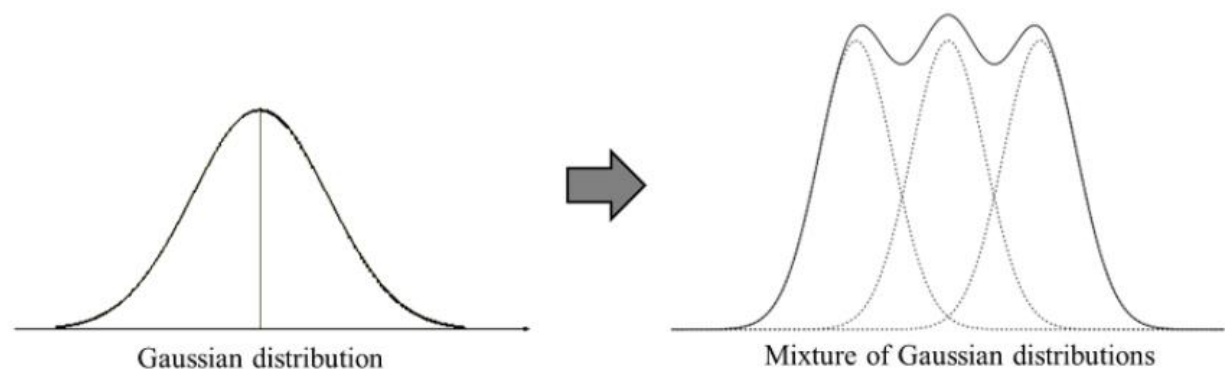
- ✓ K개의 Cluster가 존재
- ✓ x : Cluster z 안의 데이터
- ✓ 각각의 Cluster에는 선택될 확률(π), 평균(μ), 분산 (Σ) 이라는 파라미터가 존재
- ✓ 하나의 Gaussian probability density function

$$\begin{aligned} p(x, z) &= p(z)p(x|z) \\ &= \pi_z N(x|\mu_z, \Sigma_z) \end{aligned}$$

02 | Gaussian Mixture Model

Theory

- Gaussian Mixture Model (GMM)



$$\begin{aligned} p(x, z) &= p(z)p(x|z) \\ &= \pi_z N(x|\mu_z, \Sigma_z) \end{aligned}$$

$$p(x, z) = \pi_1 N(x|\mu_1, \Sigma_1) + \pi_2 N(x|\mu_2, \Sigma_2) + \pi_3 N(x|\mu_3, \Sigma_3)$$

- ✓ GMM은 Gaussian 분포가 여러 개 혼합된 Clustering 알고리즘
- ✓ 현실에 존재하는 복잡한 형태의 확률 분포를 위 그림과 같이 K개의 Gaussian distribution을 혼합하여 표현
- ✓ 이때, K는 데이터를 분석하고자 하는 사람이 직접 설정

02 | Gaussian Mixture Model

Theory

- Gaussian Mixture Model (GMM)

- ✓ 주어진 데이터 x 가 발생할 확률은 여러 Gaussian probability density function의 합으로 표현

$$p(x) = \sum_{z=1}^K \pi_z N(x|\mu_z, \Sigma_z)$$



- ✓ Mixing coefficient(π_z) : k 번째 Gaussian distribution이 선택될 확률
(단, 아래 두 조건을 모두 만족할 것)

$$0 \leq \pi_z \leq 1 \quad (1)$$

$$\sum_{z=1}^K \pi_z = 1 \quad (2)$$

03 | Learning the Gaussian Mixture Model

MLE

- What must we learn?
 - ✓ GMM은 $\theta = \{\pi, \mu, \Sigma\}$ 이 세가지 파라미터를 적절히 추정하는 것
 - ✓ θ 가 주어졌을 때, 데이터 셋을 표현할 확률을 나타내는 것이므로 이를 최대화하는 방향으로 모델의 θ 를 학습
 - ✓ 따라서 목적함수는 maximize log-likelihood

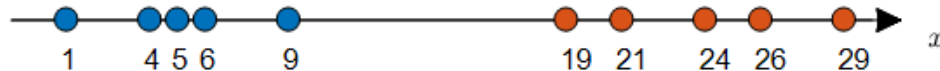
$$J(\pi, \mu, \Sigma) = \sum_{i=1}^n \log\{\sum_{z=1}^k \pi_z N(x_i | \mu_z, \Sigma_z)\}$$

- ✓ 각각의 파라미터에 대해 편 미분을 하여 파라미터 값 추정
 - i) 평균(μ_z) : 편 미분하여 0이 되는 지점 찾기
 - ii) 분산(Σ_z) : 편 미분하여 0이 되는 지점 찾기
 - iii) Mixing coefficient(π_z) : 제약 조건이 있기 때문에 라그랑주 승수로 풀기

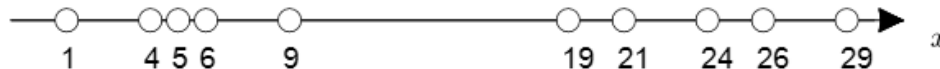
03 | Learning the Gaussian Mixture Model

MLE Problem

- Problem



✓ Label이 주어진 경우, 앞선 방법과 같이 각각에 대해 MLE를 계산하고 likelihood가 최대가 되는 파라미터 추정 가능



✓ Label이 주어지지 않은 경우, Label를 부여하기 위해서는 확률 분포가 필요 → 확률 분포를 얻기 위해서는 파라미터가 필요
→ 파라미터를 알기 위해서는 각 데이터(x)의 Label이 필요

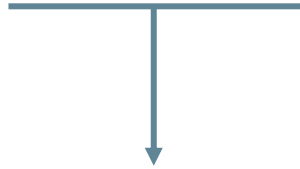
- Solution

✓ Label과 파라미터 추정을 번갈아 반복하는 방법 필요 = EM Algorithm

04 | EM Algorithm for Latent Variable Models

Key Idea

- EM Algorithm (Expectation–Maximation Algorithm)
 - ✓ 기댓값을 최대화하는 알고리즘으로 직역 가능
 - ✓ 일부 데이터의 정보가 없는 상황(Latent variable)에서 MLE를 얻어야 할 때 사용하는 알고리즘



- **Latent Variable(z)** : 실제로 관측이 되지 않았지만 관측된 데이터(x)에 상호 영향을 미치리라 판단되는 변수
본래 가지고 있는 random variable이 아닌 임의로 설정한 hidden variable을 의미

ex) Gaussian Mixture Model에서는 Label이 Latent Variable에 해당

04 | EM Algorithm for Latent Variable Models

How to work?

- EM Algorithm (Expectation–Maximation Algorithm)

1. 초기화

- ✓ 파라미터에 대한 현재의 추정 값인 θ^{old} 정의

2. Expectation-step

- ✓ θ^{old} 를 사용하여 잠재변수에 대한 Posterior을 찾고 Log-likelihood의 기댓값($Q(\theta, \theta^{old})$, Responsibility) 계산

$$Q(\theta, \theta^{old}) = \sum_z \underbrace{p(z|x, \theta^{old})}_{\text{사후확률(Posterior)}} \log p(x, z|\theta)$$

사후확률(Posterior)

04 | EM Algorithm for Latent Variable Models

How to work?

- EM Algorithm (Expectation-Maximation Algorithm)

3. Maximization-step

- ✓ E-step의 식을 최대화하는 θ 을 찾아 θ^{new} 로 설정

$$\theta^{new} = \operatorname{argmax}_{\theta} q(\theta, \theta^{old})$$

- ✓ θ 가 업데이트 되면서 log-likelihood가 증가하게 되지만 KL divergence 값이 0이 되진 않음.

두 확률분포 사이의 차이를 측정하는 지표

= 실제 분포와 추정한 분포 사이에 차이가 있다는 것을 의미 ➡ ELBO(Evidence Lower Bound) 개념 사용

04 | EM Algorithm for Latent Variable Models

How to work?

- EM Algorithm (Expectation-Maximation Algorithm)

3. Maximization-step

$$\text{증명) } \log p(x|\theta) = \log[\sum_z p(x, z|\theta)]$$

x 와 θ 에 대한

$$\text{Log-likelihood} = \log[\sum_z q(z) \left(\frac{p(x, z|\theta)}{q(z)}\right)] \quad : \text{Latent Variable에 대한 확률질량함수를 가중치로 사용하여 기댓값 계산}$$

$$\geq \underbrace{\sum_z q(z) \left(\frac{p(x, z|\theta)}{q(z)}\right)}_{\mathcal{L}(q, \theta)} \quad : \text{ELBO}(\mathcal{L}(q, \theta)) = \text{파라미터 } \theta \text{와 확률분포 } q(z) \text{에 대한 log-likelihood의 하한 값으로 해석}$$

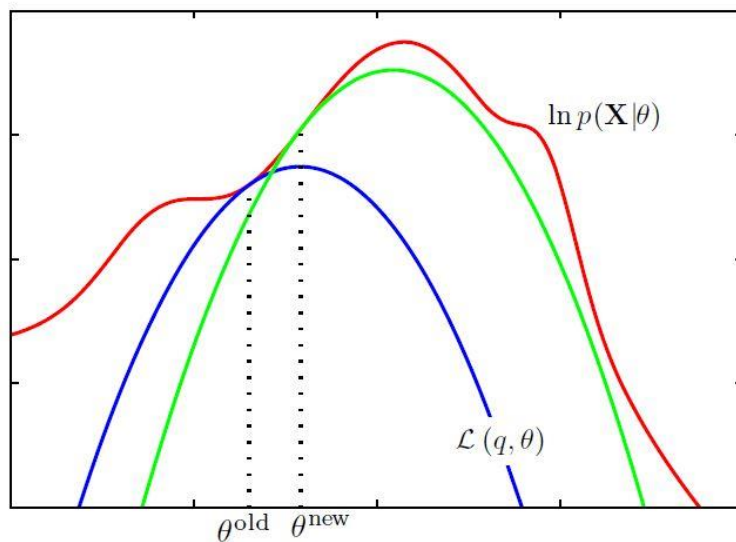
- ✓ ELBO(Evidence Lower Bound)를 최대화하는 방향으로 파라미터 θ 를 업데이트 하면 local optimum 문제 해결

04 | EM Algorithm for Latent Variable Models

How to work?

- EM Algorithm (Expectation–Maximation Algorithm)

4. Log-likelihood가 수렴하는지 확인하기



- ✓ 빨간 선 : $\ln p(X|\theta)$ = 우리가 최대화 하고 싶은 실제 분포
- ✓ 파란 선 : 초기값으로 설정한 ELBO
- ✓ 지속적으로 EM을 수행하면 초록 선과 같은 결과를 얻을 수 있게 됨.

05 | EM Algorithm for Gaussian Mixture Model

EM Algorithm

- How to learn by EM Algorithm?

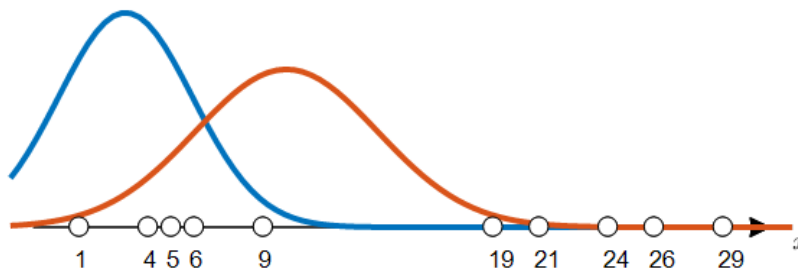
1. 모든 파라미터 $\theta = \{\pi, \mu, \Sigma\}$ 초기화 하기



- ✓ 분포가 주어진 경우, 각 데이터(x)에 대한 분포의 높이 값(likelihood)을 비교하여 Labeling 가능
- ✓ 주황색과 파란색을 각각 그룹 1, 그룹 2라고 한 경우

$$\mu_1 = 3, \sigma_1 = 2.9155$$

$$\mu_2 = 10, \sigma_2 = 3.9623$$



05 | EM Algorithm for Gaussian Mixture Model

EM Algorithm

- How to learn by EM Algorithm?

2. E-step : 초기 파라미터 $\theta = \{\pi, \mu, \Sigma\}$ 를 고정해두고 Responsibility(γ_i^j) 구하기

Random하게 부여한 분포에 기반하여 Label(z)을 찾는 과정

Latent Variable

$$\underbrace{\gamma_i^j}_{\text{Responsibility}} = P(z = j | x_i) = \frac{\pi_j N(x_i | \mu_j, \Sigma_j)}{\sum_{c=1}^K \pi_c N(x_i | \mu_c, \Sigma_c)}$$

사전확률
↑
likelihood

샘플 x_i 가 K^{th} 번째 가우시안 분포에서 발생했을 사후 확률

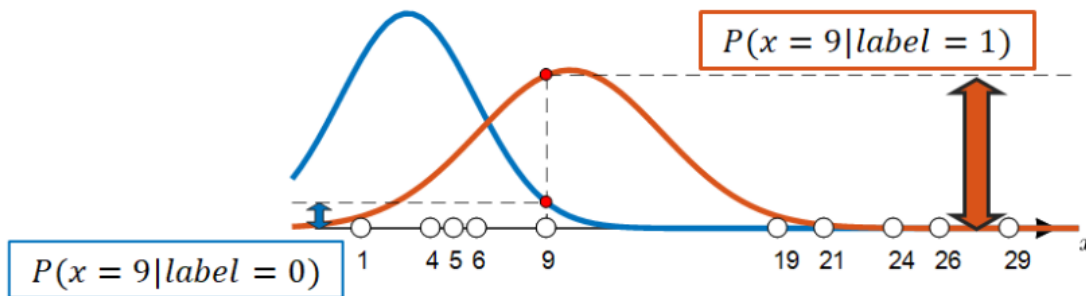
05 | EM Algorithm for Gaussian Mixture Model

EM Algorithm

- How to learn by EM Algorithm?

2. E-step : 초기 파라미터 $\theta = \{\pi, \mu, \Sigma\}$ 를 고정해두고 Responsibility(γ_i^j) 구하기

Random하게 부여한 분포에 기반하여 Label(z)을 찾는 과정



✓ 9번 데이터의 경우, likelihood를 비교했을 때 주황색 분포에 속할 가능성이 더 높음.

05 | EM Algorithm for Gaussian Mixture Model

EM Algorithm

- How to learn by EM Algorithm?

3. M-step : γ_i^j 를 활용하여 파라미터 $\theta = \{\pi, \mu, \Sigma\}$ 를 최대화 하는 과정

i) 평균(μ_z) : $\mu_c^{new} = \frac{1}{n_c} \sum_{i=1}^n \gamma_i^c x_i$

ii) 분산(Σ_z) : $\Sigma_c^{new} = \frac{1}{n_c} \sum_{i=1}^n \gamma_i^c (x_i - \mu_c^{new})(x_i - \mu_c^{new})^T$

iii) Mixing coefficient(π_z) : $\pi_c^{new} = \frac{n_c}{n}$

05 | EM Algorithm for Gaussian Mixture Model

EM Algorithm

- How to learn by EM Algorithm?

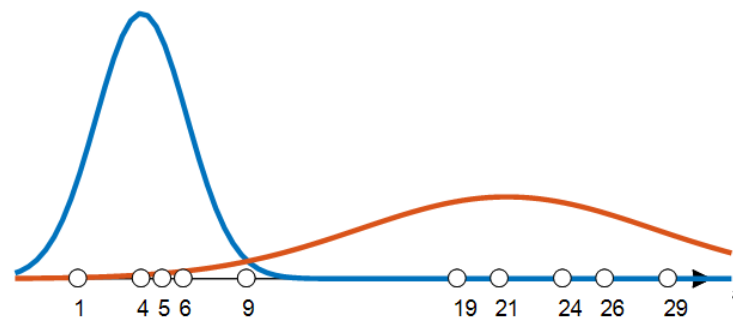
3. M-step : γ_i^j 를 활용하여 파라미터 $\theta = \{\pi, \mu, \Sigma\}$ 를 최대화 하는 과정



✓ $\mu_1 = 4, \sigma_1 = 2.1602$

$\mu_2 = 21.33, \sigma_2 = 7.0048$

✓ Random하게 가정한 분포와는 다른 분포 생성



4. log-likelihood가 수렴할 때까지 반복

06 | Gaussian Mixture Model for Clustering

EM Algorithm

- GMM for Clustering
 - ✓ K-means clustering은 각 포인트를 하나의 클러스터에만 연결하는 hard clustering method
 - ✓ 이로 인해 데이터(x)가 특정 클러스터와 얼마나 관련이 있는지에 대한 불확실성 혹은 확률이 없음

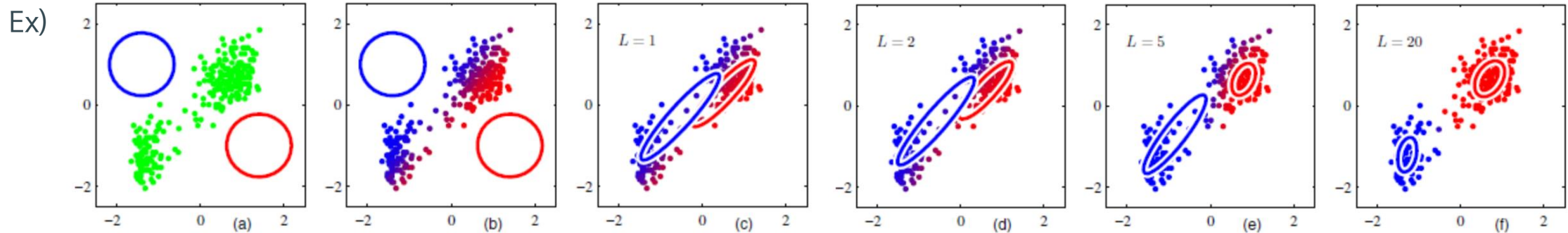


- ✓ EM Algorithm에서 구했던 각각의 Gaussian distribution의 평균 값이 cluster의 center가 되게 하면
k-means clustering처럼 clustering 가능 = GMM을 clustering이라고 부르는 이유

06 | Gaussian Mixture Model for Clustering

EM Algorithm

- GMM for Clustering



(a) : 동그라미 2개 = 두 Gaussian의 단위 표준편차 경로

초록색 점 = 확률 값(Responsibility)을 기준으로 빨강, 파랑으로 점점 칠해질 예정

(b) : E step = 확률 값(Responsibility) 계산

(c) : M step = 파라미터 θ 업데이트

(f) : 20사이클 반복 후 알고리즘 거의 수렴

Q&A

감사합니다.