Day 3 : EM algorithm

* Change of denotion

1) Expectation

Variable x's p.d.f $p(x)$, the expectation of $f(x)$ is

: $E[f(x)] \longrightarrow E_{p(x)}[f(x)] = \int f(x)p(x)\,dx$

2) PDF

$p(x;\theta) \longrightarrow p_\theta(x)$

* KL Divergence

- 두 확률분포의 차이를 줄이는 척도

- KL Divergence

for 2 prob distributions $p(x), q(x)$, KL divergence is as follows;

a) Continuous

$$D_{KL}(p\|q) = \int p(x)\log\frac{p(x)}{q(x)}\,dx$$

b) Discrete

$$D_{KL}(p\|q) = \sum_x p(x)\log\frac{p(x)}{q(x)}\,dx$$

- Properties
  - 다를 수록 값이 커짐
  - $p=q$ 이면 $D_{KL}(p\|q) = 0$ → minimum value
  - Symmetric 하지 않음. $D_{KL}(p\|q) \neq D_{KL}(q\|p)$

* KL divergence와 MLE (KL Divergence에서 어떻게 MLE식이 유도되는지?)

$p_*(x)$: 실제 모집단 분포

$D=\{x^{(1)}, x^{(2)} \cdots, x^{(N)}\}$: $p_*(x)$에서 추출한 sample

$p_\theta(x)$: θ로 조절가능한, 모집단 추정을 위해 모델링한 분포

목표는? $p_\theta(x)$를 $p_*(x)$에 최대한 가깝게 만들자

$\Leftrightarrow$ minimize $D_{KL}(p_* \| p_\theta) = \int p_*(x) \log \frac{p_*(x)}{p_\theta(x)} dx$

그러나 $p_*(x)$를 현재 알 수 없음. "몬테카를로 방법"을 통해 $D_{KL}(p_* \| p_\theta)$를 근사할 것임.

☆

$$\boxed{\begin{aligned} &\text{MLE} : \log \prod_{n=1}^{N} p_\theta(x^{(n)}) = \sum_{n=1}^{N} \log p_\theta(x^{(n)}) \\ &\text{Find } \hat{\theta} \text{ which satisfies, } \hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{n=1}^{N} \log p_\theta(x^{(n)}) \end{aligned}}$$

"Monte carlo"

- 무작위 수를 이용하여 복잡한 확률분포나 기댓값의 근사값을 구하는 방법 (정확한 p_*는 모르지만 p_*에서 나온 sample들을 가지고 있을 때, $E_{p_*(x)}[f(x)]$를 근사시키는 방법)

* $D=\{x^{(1)}, x^{(2)} \cdots x^{(N)}\}$ $(x^{(n)} \sim p_*(x))$

$E_{p_*(x)}[f(x)] = \int p_*(x) f(x) dx$ (x : continuous variable)

$\approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)}) dx$

· 기댓값 근사 방법 (with monte carlo)

$E_{p_*(x)}[f(x)] = \int p_*(x) f(x) dx$ 에서 ··· ①

1) 확률분포 $p_*(x)$에 따라 샘플 $\{x^{(1)}, x^{(2)} \cdots, x^{(N)}\}$을 추출

2) 각 $x^{(i)}$에 대해 $f(x^{(i)})$ 구하고 평균 구하기

①에서 1), 2)를 시행

$$E_{p_*(x)}[f(x)] = \int p_*(x) f(x) dx$$
$$\approx \frac{1}{N} \sum_{n=1}^{N} f(x^{(n)}) \quad (x^{(n)} \sim p_*(x))$$

"$D_{KL}(p_* \| p_\theta)$ montecarlo 근사 & MLE induction"

$D_{KL}(p_* \| p_\theta)$도 기댓값 식의 꼴로 볼 수 있음.

$$D_{KL}(p_* \| p_\theta) = \int p_*(x) \log \frac{p_*(x)}{p_\theta(x)} dx$$

$\log \frac{p_*(x)}{p_\theta(x)}$를 $f(x)$라 하면 ··· → monte carlo 근사

$$\approx \frac{1}{N} \sum_{n=1}^{N} \log \frac{p_*(x^{(n)})}{p_\theta(x^{(n)})} \quad (x^{(n)} \sim p_*(x))$$

$$\approx \frac{1}{N} \sum_{n=1}^{N} \left(\log p_*(x^{(n)}) - \log p_\theta(x^{(n)}) \right) \cdots ②$$

② 을 minimize

$$\underset{\theta}{\operatorname{argmin}}\, D_{KL}(p_* \| p_\theta) \approx \underset{\theta}{\operatorname{argmin}}\, \frac{1}{N}\sum_{n=1}^{N}\left(\cancel{\log p_*(x^{(n)})} - \log p_\theta(x^{(n)})\right)$$

$$= \underset{\theta}{\operatorname{argmin}} -\frac{1}{N}\sum_{n=1}^{N}\log p_\theta(x^{(n)})$$

$$= \underset{\theta}{\operatorname{argmin}} -\sum_{n=1}^{N}\log p_\theta(x^{(n)})$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{n=1}^{N}\log p_\theta(x^{(n)})$$

$$\therefore\ \underset{\theta}{\operatorname{argmin}}\, D_{KL}(p_* \| p_\theta) \approx \underbrace{\underset{\theta}{\operatorname{argmax}} \sum_{n=1}^{N}\log p_\theta(x^{(n)})}_{L_{MLE}}$$

* Review : $p(x)$ for model containing latent variable

$p(x)$가 Σ를 담아내서 MLE 식이 $\Sigma \log \Sigma$ log-sum 형태를 가짐 ⇒ 해석적으로 $\underset{\theta}{\operatorname{argmax}}\, L(D;\theta)$를 만족하는 θ를 해석적으로 찾아낼 수 없음

So, we use EM algorithm to find θ maximizing $L(D;\theta)$

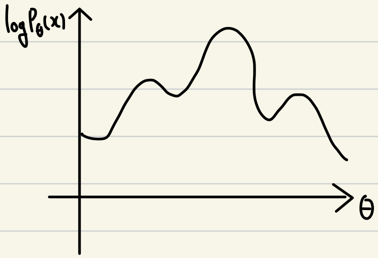
$$\left(L(D;\theta) = \sum_{n=1}^{N}\log p(x^{(n)})\right)$$

Generalization of EM for model with latent variable Z

* EM algorithm for "Single" data x.

E - M

(expectation) (Maximize)



이렇게 $\log P_\theta(x)$가 복잡한 이유는 $P_\theta(x)$가 $\sum_z$를 포함하기 때문.

$\log P_\theta(x) = \log \sum_z P_\theta(x, z)$ 에서 $\sum_z$를 없애주는 것이 목표!

$\log P_\theta(x) = \log \frac{P_\theta(x,z)}{P_\theta(z|x)}$ (곱셈정리 $p(x,y) = P(x|y)P(y)$) $\Rightarrow \log P_\theta(x)$ 변형

$\log \frac{P_\theta(x,z)}{P_\theta(z|x)}$ 에서 $P_\theta(z|x) = \frac{P_\theta(z,x)}{P_\theta(x)} = \frac{P_\theta(z,x)}{\sum_z P_\theta(x,z)}$ 로 여전히 $\sum_z$를 가지고 있음..., log-sum을 취하고 있음.

Solution1: $P_\theta(z|x)$의 근사분포로 임의의 확률분포 $q(z)$를 이용.

$$\log P_\theta(x) = \log \frac{P_\theta(x,z)}{P_\theta(z|x)} \times \underbrace{\frac{q(z)}{q(z)}}_{1}$$

$$= \log \frac{P_\theta(x,z)}{q(z)} \times \frac{q(z)}{P_\theta(z|x)}$$

$$= \log \frac{P_\theta(x,z)}{q(z)} + \log \frac{q(z)}{P_\theta(z|x)}$$

① $P_\theta(z|x)$를 $q(z)$로 바꿈. (ok)

② → 아직 $P_\theta(z|x)$가 남아있음 (문제)

Solution 2: ②번 항을 KL 식으로 바꿔주기

$\log P_\theta(x)$ → $\sum_z q(z) = 1.$

$= \log P_\theta(x) \sum_z q(z)$

$= \sum_z q(z) \log P_\theta(x)$ → sol1에서 도출된 식

$= \sum_z q(z) \left(\log \frac{P_\theta(x,z)}{q(z)} + \log \frac{q(z)}{P_\theta(z|x)} \right)$ → KL로 유도됨!

$= \sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)} + \sum_z q(z) \log \frac{q(z)}{P_\theta(z|x)}$

$= \underset{①}{\sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)}} + \underset{②}{D_{KL}(q(z) \,||\, P_\theta(z|x))}$

⇒ ①,② 어디에도 log − Σ 형태가 없음!

∴ for single data x,

$$\log P_\theta(x) = \sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)} + D_{KL}(q(z) \,||\, P_\theta(z|x))$$

(latent variable z가 있는 모델에 대해 $\log P_\theta(x)$를 log − $\sum_z$이 없는 형태로 두 항으로 쪼개어 변형시킴)

* ELBO

$D_{KL} \geq 0$ 임을 이용하면,

$\log P_\theta(x) \geq \underset{①}{\sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)}}$ 가 성립함.

①는 항상 $\log P_\theta(x)$의 하한임. (증가하한)

이때, ①는 ELBO항이라고함.

$$\mathrm{ELBO}(x; q, \theta) = \sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)}$$

① $\mathrm{ELBO}(x; q, \theta) \leq \log P_\theta(x)$

② Sum-log form : 해석적인 최적화 가능.

⇒ ①,②에 의해 $\mathrm{ELBO}(x; q, \theta)$를 maximize하면서 간접적으로 $\log P_\theta(x)$를 키운다. 이것이 E-M algorithm

✳ ELBO$(x; q, \theta)$ optimization : E-M

q, θ를 한꺼번에 최적화하는 것이 어려움. ⇒ 하나하나씩 갱신.

- E : θ fix , $q(z)$ update
- M : θ update , $q(z)$ fix

E : 간접적인 ELBO를 $\log P_\theta(x)$에 맞춰주기. (θ fix , $q(z)$ update)

$$\log P_\theta(x) = \underset{①}{\text{ELBO}(x; q, \theta)} + \underset{②}{D_{KL}(q(z) \,||\, P_\theta(z|x))}$$

①+②는 항상 고정 ②를 0으로 만들면 ELBO$(x; q, \theta) = \log P_\theta(x)$를 만족함

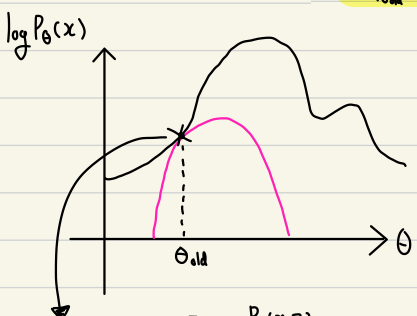
How?

for $\theta = \theta_{old}$ (fix)

$q(z) = P_{\theta_{old}}(z|x)$를 만족하도록 q를 update.

$$\begin{aligned}\text{ELBO}(x; q = P_{\theta_{old}}(z|x), \theta) &= \sum_z q(z) \log \frac{P_\theta(x,z)}{q(z)} \\ &= \sum_z P_{\theta_{old}}(z|x) \log \frac{P_\theta(x,z)}{P_{\theta_{old}}(z|x)} \\ &= E_{P_{\theta_{old}}(z|x)}\left[\log \frac{P_\theta(x,z)}{P_{\theta_{old}}(z|x)}\right]\end{aligned}$$

↳ $q(z) = P_{\theta_{old}}(z|x)$ 일때 기댓값으로 표현되어서 E-step이라 불림.



$E_{P_{\theta_{old}}(z|x)}\left[\log \frac{P_\theta(x,z)}{P_{\theta_{old}}(z|x)}\right] = \log P_\theta(x)$ 인 지점. (θ_{old} 고정, q update를 통해)

↳ Expectation 형식으로 표현된 ELBO

M : maximize ELBO using θ.

① q fix, θ update

② $\arg\max_{\theta} \mathrm{ELBO}(x; q, \theta)$

$$\frac{1}{\partial\theta}\mathrm{ELBO}(x; q = P_{\theta_{old}}(z|x), \theta) = \frac{1}{\partial\theta} E_{P_{\theta_{old}}(z|x)}\left[\log \frac{P_\theta(x,z)}{P_{\theta_{old}}(z|x)}\right] = 0$$

앞의 E-step으로 구한 식에 대해 θ 편미분 후 0이 되는 지점 찾아서 maximize

☆ 정리

E : $q(z)$ update → $\mathrm{ELBO}(x; q, \theta) = \log P_\theta(x)$

M : θ update → ELBO 최대화를 통해 간접적으로 $\log P_\theta(x)$ 높이기 ($\log P_\theta(x) \geq \mathrm{ELBO}(x; q, \theta)$)

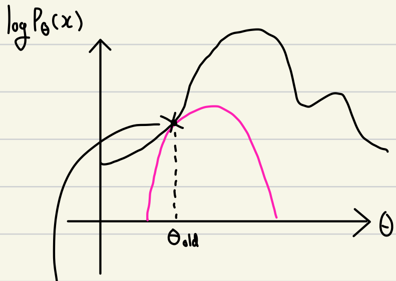
Property)

"$\log p(x; \theta_{new}) \geq \log p(x; \theta_{old})$"

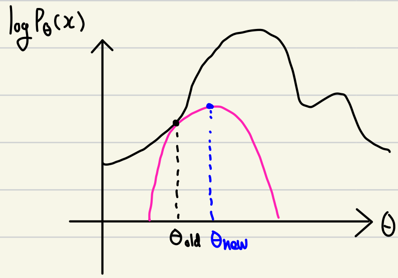
- With EM update, the log likelihood $P_\theta(x)$ is monotonously increasing.
- Iteratively conduct EM until the increase of log-likelihood is below the threshold. (or the updated amount of parameter is below the threshold)

※ ELBO in Graph

① E



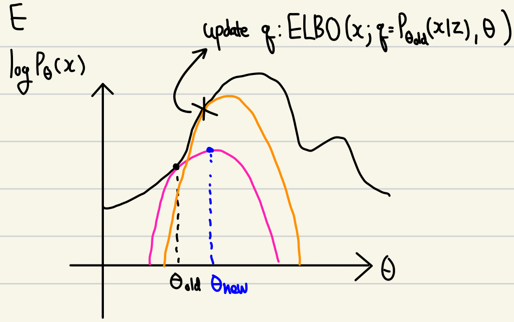
② M



$E_{P_{\theta_{old}}(z|x)}\left[\log \frac{P_\theta(x,z)}{P_{\theta_{old}}(z|x)}\right] = \log P_\theta(x)$ 인 지점. (θ_{old} 고정, q update를 통해)

ㄴ Expectation 형식으로 표현된 ELBO

③ E



"Iterate E-M!"

* EM Algorithm for "Multiple datas"

$D = \{ x^{(1)}, x^{(2)} \cdots x^{(N)} \}$ (Sample from model that has latent variable)

$Q = \{ q^{(1)}, q^{(2)} \cdots q^{(N)} \} \Rightarrow q(z)$ corresponding to each data $x^{(i)}$

latent variable z corresponding to n^{th} data $x^{(n)}$

$$\mathcal{L}(D) = \sum_{n=1}^{N} \log \sum_{z^{(n)}} P_\theta(x^{(n)}, z^{(n)}) \geq \sum_{n=1}^{N} \mathrm{ELBO}(x^{(n)}; q^{(n)}, \theta)$$

$$= \sum_{n=1}^{N} \sum_{z^{(n)}} q^{(n)}(z^{(n)}) \log \frac{P_\theta(x^{(n)}, z^{(n)})}{q^{(n)}(z^{(n)})}$$

($q^{(n)}(z^{(n)})$: $z^{(n)}$ latent variable's distribution)

Single data x 에 대한 E-M에서 E단계의 최적 $q(z)$는 $q(z) = P_\theta(z|x)$ (KL항 0으로 → $\mathrm{ELBO} = \log P_\theta(x)$)

↓

Multiple이라면 $q^{(n)}(z^{(n)}) = P_\theta(z^{(n)} | x^{(n)}) \Rightarrow$ N개의 equation 나옴.

데이터 N개에 대한 E-M 종료 판단? mean log-likelihood의 증가량.

$\Rightarrow \frac{1}{N} \sum_{n=1}^{N} \log P_\theta(x^{(n)}; \theta)$

* 정리

1) E: fix θ, update each $q^{(n)}(z^{(n)}) = P_\theta(z^{(n)} | x^{(n)})$

2) M: fix $\{ q^{(1)}(z^{(1)}), \cdots q^{(n)}(z^{(n)}) \}$, update θ : $\underset{\theta}{\arg\max} \sum_{n=1}^{N} \mathrm{ELBO}(x^{(n)}; q^{(n)}, \theta)$

1), 2) iterate until the mean log-likelihood increase is below the threshold

$\left(\frac{1}{N} \sum_{n=1}^{N} P_\theta(x^{(n)}) \right)$

EM for Gaussian Mixture Model (GMM)

* GMM review

GMM
- x : 관측변수
- z : 잠재변수 (z는 1~K까지의 이산값을 취하는 확률변수)
- θ : 매개변수 $\theta = \{\phi, u, \Sigma\}$,
  - $\phi = \{\phi_1, \phi_2 \cdots \phi_K\}$ prob of selecting K^{th} multivariate Gaussian Dist.
  - $u = \{u_1, u_2 \cdots u_K\}$ mean vector of K^{th} MGD.
  - $\Sigma = \{\Sigma_1, \Sigma_2 \cdots \Sigma_K\}$ covariance matrix of K^{th} MGD.

$\log P_\theta(x)$ for GMM

$$p(x;\theta) = \sum_{j=1}^{K} p(x, z=j;\theta) \quad \text{(marginalization)}$$

$$= \sum_{j=1}^{K} p(z=j;\theta)\, p(x \mid z=j;\theta)$$

$$= \sum_{j=1}^{K} \phi_j \, N(x; u_j, \Sigma_j)$$

※ EM

$D = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ sampled from GMM : Sample

$\frac{1}{N}\sum_{n=1}^{N} \log P_\theta(x^{(n)}) = \frac{1}{N}\sum_{n=1}^{N} \log \sum_{j=1}^{k} \phi_j N(x^{(n)}; u_j, \Sigma_j)$: mean log-likelihood, a criteria for termination of EM

- E : θ fix, $Q = \{q_1, q_2 \cdots q_n\}$ update. (ELBO approximation to $\log P_\theta(x^{(n)})$)

KL=0을 위한 갱신식

$$q^{(n)}(z^{(n)}=k) = p_\theta(z^{(n)}=k \mid x^{(n)})$$

($q^{(n)}(z^{(n)}) = p_\theta(z^{(n)} \mid x^{(n)})$ 에서 $z=k$만 특정)

$$= \frac{P_\theta(x^{(n)}, z^{(n)}=k)}{P_\theta(x^{(n)})}$$

marginalize) $\sum_j P_\theta(x^{(n)}, z^{(j)})$

$$= \frac{\phi_k N(x^{(n)}; u_k, \Sigma_k)}{\sum_j \phi_j N(x^{(n)}; u_j, \Sigma_j)}$$

abbreviation : $q^{(n)}(z^{(n)}=k) \longrightarrow q^{(n)}(k)$

$\therefore$ for all n, k (for all possible n datas and all possible latent variables)

renew $q^{(n)}(z^{(n)}=k)$ to ...

$$q^{(n)}(z^{(n)}=k)\,(= q^{(n)}(k)) = \frac{\phi_k N(x^{(n)}; u_k, \Sigma_k)}{\sum_j \phi_j N(x^{(n)}; u_j, \Sigma_j)}$$

↳ Z가 discrete하다면, n×k개의 갱신식이 생긴다...?

• M : $Q = \{ q^{(1)}(z^{(1)}), q^{(2)}(z^{(2)}) \cdots q^{(N)}(z^{(N)}) \}$ fix, $\theta = \{ \phi, \mu, \Sigma \}$ update. (ELBO maximize)

ELBO for $D = \{ x^{(1)}, x^{(2)} \cdots x^{(N)} \}$

$$\sum_{n=1}^{N} \mathrm{ELBO}(x^{(n)}; q^{(n)}, \theta)$$

$$= \sum_{n=1}^{N} \sum_{j=1}^{K} \underbrace{q^{(n)}(z^{(n)}=j)}_{= q^{(n)}(j)} \log \frac{P_\theta(x^{(n)}, z^{(n)}=j)}{q^{(n)}(z^{(n)}=j)}$$

$$= \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \log \frac{\phi_j N(x^{(n)}; \mu_j, \Sigma_j)}{q^{(n)}(j)}$$

$$= \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \log \phi_j N(x^{(n)}; \mu_j, \Sigma_j) - \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \log q^{(n)}(j)$$

not containing $\theta = \{ \phi, \mu, \Sigma \}$: ignore

$\therefore$ The objective function is $J(\phi, \mu, \Sigma) = \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \log \phi_j N(x^{(n)}; \mu_j, \Sigma_j)$

$$= \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \left(\log \phi_j + \log N(x^{(n)}; \mu_j, \Sigma_j) \right)$$

Purpose : find θ which satisfies $\underset{\theta}{\operatorname{argmax}}\, J(\theta)$, where $\theta = \{ \phi, \mu, \Sigma \}$

How? : find $\frac{\partial J}{\partial \mu} = 0$, $\frac{\partial J}{\partial \Sigma} = 0$, $\frac{\partial J}{\partial \phi} = 0$ and update μ, Σ, ϕ stepwisely.

Where,

$\mu = \{ \mu_1, \mu_2 \cdots \mu_K \}$, $\Sigma = \{ \Sigma_1, \Sigma_2 \cdots \Sigma_K \}$, $\phi = \{ \phi_1, \phi_2 \cdots, \phi_K \}$

1) $\dfrac{\partial J}{\partial u}$ $\left(u = \{u_1, u_2 \cdots u_K\} \right)$

$$\frac{\partial J}{\partial u_k} = \frac{\partial}{\partial u_k} \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \left(\cancel{\log \phi_j} + \log N(x^{(n)}; u_j, \Sigma_j) \right)$$

$$= \frac{\partial}{\partial u_k} \sum_{n=1}^{N} \sum_{j=1}^{K} q^{(n)}(j) \log N(x^{(n)}; u_j, \Sigma_j)$$

$$= \sum_{n=1}^{N} q^{(n)}(k) \underbrace{\frac{\partial}{\partial u_k} \log N(x^{(n)}; u_k, \Sigma_k)}_{\text{다변량 정규분포의 평균 벡터에 대한 미분.}}$$

$$* \quad \frac{\partial}{\partial u} \log N(x; u, \Sigma) = \Sigma^{-1}(x-u)$$

$$= \sum_{n=1}^{N} q^{(n)}(k) \Sigma_k^{-1} (x^{(n)} - u_k)$$

$$\therefore \frac{\partial J}{\partial u_k} = \sum_{n=1}^{N} q^{(n)}(k) \Sigma_k^{-1} (x^{(n)} - u_k)$$

$$\frac{\partial J}{\partial u_k} = 0 \iff \sum_{n=1}^{N} q^{(n)}(k) \Sigma_k^{-1} (x^{(n)} - u_k) = 0$$

$$\iff \Sigma_k \sum_{n=1}^{N} q^{(n)}(k) \Sigma_k^{-1} (x^{(n)} - u_k) = 0 \times \Sigma_k$$

$$\iff \sum_{n=1}^{N} q^{(n)}(k) \cancel{\Sigma_k \Sigma_k^{-1}} (x^{(n)} - u_k) = 0 \quad (q^{(n)}(k) \text{ is a Scalar})$$

$$\iff \sum_{n=1}^{N} q^{(n)}(k) (x^{(n)} - u_k) = 0$$

$$\iff \sum_{n=1}^{N} q^{(n)}(k)\, u_k = \sum_{n=1}^{N} q^{(n)}(k)\, x^{(n)}$$

$$\iff u_k \sum_{n=1}^{N} q^{(n)}(k) = \sum_{n=1}^{N} q^{(n)}(k)\, x^{(n)}$$

$$\iff u_k = \frac{\sum_{n=1}^{N} q^{(n)}(k)\, x^{(n)}}{\sum_{n=1}^{N} q^{(n)}(k)}$$

$\therefore$ u_k의 갱신식 $\quad u_k = \dfrac{\sum_{n=1}^{N} q^{(n)}(k)\, x^{(n)}}{\sum_{n=1}^{N} q^{(n)}(k)}$ $\Rightarrow$ 이를 $u_1, u_2 \cdots u_K$에 대해 반복.

2) $\dfrac{\partial J}{\partial \Sigma}$ $\left(\Sigma = \{\Sigma_1, \Sigma_2 \cdots \Sigma_K\}\right)$

[abbreviate mid - process]

$$\frac{\partial J}{\partial \Sigma_k} = 0 \iff \Sigma_k = \frac{\sum_{n=1}^{N} q^{(n)}(k)\,(x^{(n)} - u_k)(x^{(n)} - u_k)^T}{\sum_{n=1}^{N} q^{(n)}(k)} \quad \cdots ①$$

(이때 u_k는 $\dfrac{\partial L}{\partial u_k} = 0$ 에 대해 갱신된 u_k)

①의 과정을 $\Sigma_1, \Sigma_2 \cdots \Sigma_k$에 대해 진행.

✳ Lagrange Multiplier Method

제약조건이 존재할 때의 최적화 문제 : Lagrange Multiplier Method

$f(x,y)$ optimization under constraint $g(x,y)=0$ constraint와 관련된 식

main idea) objective function인 f와 constraint function g의 기울기가 같은 방향일 때 목적함수 $f(x,y)$에 대한 최적해 존재

$\therefore \nabla f(x,y) = \lambda \nabla g(x,y)$ 일 때 $f(x,y)$에 대한 최적해 존재

lagrange multiplier

Let $L(x,y,\lambda) : f(x,y) - \lambda g(x,y)$

lagrangian

(tip) 제약조건이 여러개? λ도 여러개 $\{\lambda_1, g_1(x,y)\}$
$\{\lambda_2, g_2(x,y)\}$

• 최적해 찾는 방법

① lagrangian 함수 정의

② 목적함수의 변수 + λ에 대해 편미분 후 0이 되는 식들에 대해 연립방정식 풀이.

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial \lambda} = 0$$

ex) Minimize $f(x,y) = x^2+y^2$, where constrant $x+y=1$.

① Lagrange function

$$L(x,y,\lambda) = f(x,y) - \lambda g(x,y)$$
$$= x^2+y^2-\lambda(x+y-1)$$

$\hookrightarrow g(x,y)=0$ 이므로!

② 편미분

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = 0, \quad \frac{\partial L}{\partial \lambda} = 0$$

$2x - \lambda = 0 \cdots ①$

$2y - \lambda = 0 \cdots ②$

$-(x+y-1) = 0 \cdots ③$

$2x = \lambda, \ 2y = \lambda, \ x+y=1$

$x = y$ (①,②)

$x = \frac{1}{2}, y = \frac{1}{2}, \lambda = 1$ ③

$\therefore f(\frac{1}{2}, \frac{1}{2})$일 때 최소값을 가짐!

3) $\dfrac{\partial J}{\partial \phi}$ $\left(\phi = \{\phi_1, \phi_2 \cdots \phi_K\} \right)$

$J(\phi, u, \Sigma) = \sum_{n=1}^{N}\sum_{j=1}^{K} q^{(n)}(j)\left(\log \phi_j + \log N(x^{(n)}; \mu_j, \Sigma_j)\right)$ 에서 ϕ에 관련된 항은 아래와 같다.

$\sum_{n=1}^{N}\sum_{j=1}^{K} q^{(n)}(j) \log \phi_j$ (ϕ에 대해 미분에 관여하는 항) ··· Ⓐ

But, ϕ에 대한 $J(\phi, \mu, \Sigma)$ 최적화는 constraint가 있음.

constraint : $\sum_{j=1}^{K} \phi_j = 1$

① objective function $f(\phi) = \sum_{n=1}^{N}\sum_{j=1}^{K} q^{(n)}(j) \log \phi_j$

② constraint function $g(\phi) = \sum_{j=1}^{K} \phi_j - 1$

$L : f(\phi) - \lambda g(\phi)$

lagrange multiplier λ를 β로 수정.

$$L = \sum_{n=1}^{N}\sum_{j=1}^{K} q^{(n)}(j) \log \phi_j + \beta\left(\sum_{j=1}^{K}\phi_j - 1\right)$$

1) $\dfrac{\partial L}{\partial \phi_k} = \sum_{n=1}^{N} q^{(n)}(k) \dfrac{\partial}{\partial \phi_k} \log \phi_k + \beta$ $\left(\dfrac{\partial L}{\partial \phi_1} = 0, \dfrac{\partial L}{\partial \phi_2} = 0 \cdots \dfrac{\partial L}{\partial \phi_k} = 0 \text{ 에 대해} \ldots \right)$

$= \sum_{n=1}^{N} \dfrac{q^{(n)}(k)}{\phi_k} + \beta$

2) $\dfrac{\partial L}{\partial \beta} = \sum_{j=1}^{K} \phi_j - 1$

$\dfrac{\partial L}{\partial \beta} = 0 \iff \sum_{j=1}^{K} \phi_j = 1$

$\therefore \dfrac{\partial L}{\partial \phi_k} = 0 \iff \phi_k = \dfrac{\sum_{n=1}^{N} q^{(n)}(k)}{-\beta}$ ··· ①

⇒ 총 k+1개의 방정식 : $\dfrac{\partial L}{\partial \phi_1} = 0$, $\dfrac{\partial L}{\partial \phi_2} = 0$ $\cdots$ $\dfrac{\partial L}{\partial \phi_k} = 0$, $\dfrac{\partial L}{\partial \beta} = 0$

· 여러 나랍 : Φ_k를 2)에 대입
↓
$\phi_1, \phi_2 \cdots \phi_k$에 대해 $\frac{\partial L}{\partial \phi_j} = 0$을 푼 식

$$\sum_{j=1}^{K} \phi_k = 1$$

$$\Leftrightarrow \sum_{j=1}^{K} \frac{\Sigma_{n=1}^{N} q^{(n)}(j)}{-\beta} = 1$$

$$\Leftrightarrow -\beta = \sum_{j=1}^{K}\sum_{n=1}^{N} q^{(n)}(j)$$

$$\Leftrightarrow -\beta = \sum_{n=1}^{N}\underbrace{\sum_{j=1}^{K} q^{(n)}(j)}_{1}$$

$$\Leftrightarrow -\beta = N$$

$-\beta = N$을 ①에 대입하면 Ⓐ를 최대화하는 Φ_k를 구할 수 있음.

$$\therefore \Phi_k = \frac{\Sigma_{n=1}^{N} q^{(n)}(k)}{N}$$

* EM in GMM summary

① E

for n, k, $q^{(n)}(k) = \frac{\phi_k N(x^{(n)}; u_k, \Sigma_k)}{\Sigma_{j=1}^{K} \phi_j N(x^{(n)}; u_j, \Sigma_k)}$

② M

for each k,

$$\phi_k = \frac{1}{N}\sum_{n=1}^{N} q^{(n)}(k)$$

$$u_k = \frac{\sum_{n=1}^{N} q^{(n)}(k) x^{(n)}}{\sum_{n=1}^{N} q^{(n)}(k)}$$

$$\Sigma_k = \frac{\Sigma_{n=1}^{N} q^{(n)}(k)(x^{(n)} - u_k)(x^{(n)} - u_k)^T}{\Sigma_{n=1}^{N} q^{(n)}(k)}$$

③ Iterate ①,② step until mean-log-likelihood increase amount is below threshold!

$$\text{Mean log-likelihood} = \frac{1}{N}\sum_{n=1}^{N} \log \sum_{j=1}^{K} \phi_j N(x^{(n)}, u_j, \Sigma_j)$$