<Multinomial Distribution>

$K=6 \rightarrow x_i = (0,0,1,0,0,0)$ 2번째 option selected

$\text{Sum}(x_i)=1$
(하나만 선택)

$$P(X|\mu) = \prod_{k=1}^{K} \mu_k^{x_k}$$

μ_k: K번째를 선택할 확률
$(\mu_k \geq 0 \quad \sum_k \mu_k = 1)$

- Given Dataset D with N selections $x_1\ x_2 \cdots x_N$

$$P(X|\mu) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mu_k^{x_{nk}} = \prod_{k=1}^{K} \mu_k^{\sum_{n=1}^{N} x_{nk}} = \prod_{k=1}^{K} \mu_k^{m_k}$$

$$m_k = \sum_{n=1}^{N} x_{nk}$$

maximum likelihood solution of μ

① Maximize $P(X|\mu) = \prod_{k=1}^{K} \mu_k^{m_k}$

② 제약 조건 $\mu_k \geq 0, \sum_k \mu_k = 1$

↓

라그랑주 사용

Maximize $f(x,y)$
Subject to $g(x,y)=c$

$$L(x,y,\lambda) = f(x,y) + \lambda(g(x,y)-c)$$

$$L(\mu, m, \lambda) = \sum_{k=1}^{K} m_k \cdot \ln \mu_k + \lambda\left(\sum_{k=1}^{K} \mu_k - 1\right)$$

loglikelihood로 계산용이

$$\frac{d}{d\mu_k} L(\mu, m, \lambda) = \frac{m_k}{\mu_k} + \lambda = 0 \qquad \therefore \mu_k = \frac{-m_k}{\lambda}$$

$\lambda = ?$

$$\sum_k \mu_k = 1 = \sum_k \frac{-m_k}{\lambda}$$

$$-\lambda = \sum_k m_k \qquad m_k = \sum_{n=1}^{N} x_{nk}$$

$$\therefore -\lambda = \sum_k^K \sum_{n=1}^{N} x_{nk} = N \quad (\text{데이터 수})$$

$$\mu_k = \frac{m_k}{N}$$

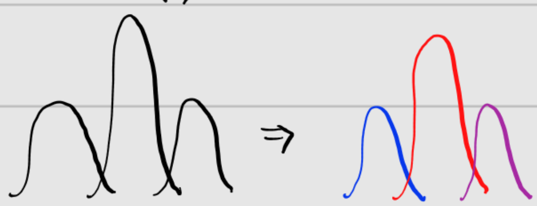
Multivariate Gaussian Distribution

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \qquad \Sigma = \begin{pmatrix} V(x_1) & cov(x_1, x_2) & \cdots & cov(x_1, x_n) \\ & V(x_2) & & \\ & & \ddots & \\ & & & V(x_n) \end{pmatrix}$$

$$\hat{\mu} = \frac{\sum_{n=1}^{N} x_n}{N} \qquad \hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{\mu})(x_n - \hat{\mu})^T$$

<Mixture Model>



- subpopulation
- 기존 정규분포만으로 설명X
- 3개의 정규분포를 mix

⇒ Mixture distribution

weight 처럼 생각

$$P(x) = \sum_{k=1}^{K} \pi_k \, N(x ; \mu, \sigma^2)$$

→ k개의 normal distribution mixing

$$P(x) = \sum_{k=1}^{K} P(z_k) \cdot P(x|z)$$

$z_k \in \{0,1\}$, $\sum_k z_k = 1$ $\quad P(z_k=1) = \pi_k$

π_k : k번째 Gaussian Distribution 선택될 확률

$0 \le \pi_k \le 1$, $\sum_{k=1}^{K} \pi_k = 1$

$$P(x|z_k=1) = N(x|\mu_k, \Sigma_k)$$

$$r(z_{nk}) = p(z_{nk}=1 | x_n)$$ $r \rightarrow$ responsibility

→ 주어진 데이터 x_n에 대해 어떤 Gaussian distribution에서 생성?

$z_{nk} \in \{0,1\}$ ⇒ $z_{nk}=1$ 이라면 x_n은 k번째 Gaussian distribution

∴ K개의 $r(z_{nk})$ 구하고, 가장 큰값 선택

$$r(z_{nk}) = P(z_{nk}=1 | x_n) = \frac{P(z_{nk}=1) \cdot P(x_n | z_{nk}=1)}{\sum_{j=1}^{K} p(z_{nj}=1) \cdot P(x|z)}$$

$$= \frac{\pi_k \cdot N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot N(x|\mu_j, \Sigma_j)}$$

Maximization

$$\ln P(X|\pi,\mu,\Sigma) = \ln \prod_{n=1}^{N} P(x_n|\pi,\mu,\Sigma)$$

$$= \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(x_n|\mu_k,\Sigma_k) \right\}$$

$$\frac{d}{d\mu_k} \ln P(X|\pi,\mu,\Sigma) = \sum_{n=1}^{N} r(z_{nk})(x_n-\mu_k) = 0$$

$$\therefore \mu_k = \frac{\sum_{n=1}^{N} r(z_{nk})\cdot x_n}{\sum_{n=1}^{N} r(z_{nk})}$$

$$\frac{d}{d\Sigma_k} \ln P(X|\pi,\mu,\Sigma) = 0$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} r(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N} r(z_{nk})}$$

$$\frac{d}{d\pi_k} \ln P(X|\pi,\mu,\Sigma) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) = 0$$

$$\pi_k = \frac{\sum_{n=1}^{N} r(z_{nk})}{N}$$

$\{X, Z\}$: complete set of variables

X: Observed variables

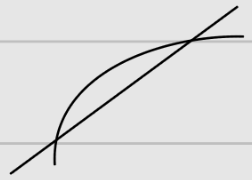
Z: hidden variables

θ: parameters for distribution

$$P(X|\theta) = \sum_{Z} P(X, Z|\theta)$$

$$\ln P(X|\theta) = \ln\left\{\sum_{Z} P(X, Z|\theta)\right\} = \ln\left\{\sum_{Z} q(Z)\frac{P(X, Z|\theta)}{q(Z)}\right\}$$

Jensen's enquality 활용



ln → Concave

$$\ln\left\{\sum_{Z} q(Z)\frac{P(X, Z|\theta)}{q(Z)}\right\} \geq \sum_{Z} q(Z)\ln\frac{P(X, Z|\theta)}{q(Z)}$$

$$\Downarrow$$

$$\sum_{Z}\left(q(Z)\ln P(X, Z|\theta) - q(Z)\ln q(Z)\right)$$

entrophy?

$$H(X) = -\sum_{x} P(X=x)\cdot\log_b P(X=x)$$

① $E_{q(Z)} \ln P(X, Z|\theta) - H(q)$ 를 Maximize 하기

⇒ 우리의 원래 식 $\ln\left\{\sum_{Z} q(Z)\frac{P(X, Z|\theta)}{q(Z)}\right\}$ 의 low boundary↑

$$\sum_z q(z) \ln \frac{P(x,z|\theta)}{q(z)} = \sum_z q(z) \ln \frac{P(z|\theta,x)\cdot P(x|\theta)}{q(z)}$$

$$= \sum_z \left\{ q(z) \ln \frac{P(z|x,\theta)}{q(z)} + \underbrace{q(z)}_{\sum_z q(z)=1} \ln P(x|\theta) \right\}$$

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

$$= \ln P(x|\theta) - \sum_z q(z) \ln \frac{q(z)}{P(z|x,\theta)}$$

②

초기 $\ln P(x|\theta) \geq$

기존 $\ln P(x|\theta)$ 에서 inequality를 만드는 factor 찾음

KL-Divergence 활용

$$KL(P\|Q) = \sum_i P(i) \ln\left(\frac{P(i)}{Q(i)}\right)$$

$$\Rightarrow KL(q(z) \,\|\, P(z|x,\theta))$$

$\ln P(X|\theta) \geq ①$ $\qquad \ln P(X|\theta) \geq ②$

$① \rightarrow E_{q(z)} \ln P(X,Z|\theta) + H(q) = Q(\theta, q)$

$② \rightarrow \ln P(X|\theta) - \sum_z \left\{ q(z) \ln \frac{q(z)}{P(z|X,\theta)} \right\} = L(\theta, q)$

② 식의 $\ln P(X|\theta)$ 고정, 특정시간 θ^t

$q^t(z) = p(z|X,\theta^t) \rightarrow$ KL divergence $= 0$

θ 바뀌면 다시 KL-d ≠ 0

① 식에 $q^t(z) = p(z|X,\theta^t)$ 대입

$$Q(\theta, q^t) = E_{q^t(z)} \ln P(X,Z|\theta^t) + H(q^t)$$

다시 KL-d=0

$$\theta^{t+1} = \operatorname{argmax} Q(\theta, q^t) = \operatorname{argmax} E_{q^t(z)} \ln P(X,Z|\theta)$$

Q. $K=3$ GMM Model $P(Z_2|X=2.5)$

$P(Z_1) = 0.5$, $P(Z_2) = 0.4$, $P(Z_3) = 0.1$

	C1	C2	C3
	$m=1$	$m=2$	$m=3$
	$\sigma=1$	$\sigma=3$	$\sigma=0.5$

$P(X=2.5|Z_1) = 0.129$

$P(X=2.5|Z_2) = 0.1311$

$P(X=2.5|Z_3) = 0.4839$

$$\frac{0.1311 \times 0.4}{0.129 \times 0.5 + 0.1311 \times 0.4 + 0.4839 \times 0.1} = 0.3167$$