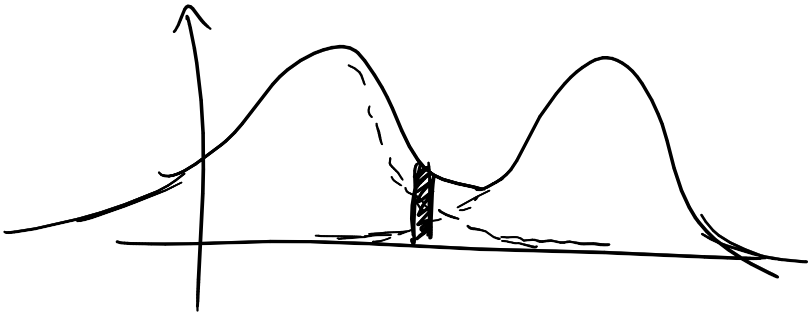
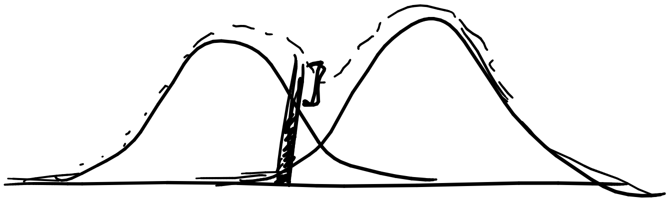
1. GMM : Gaussian Mixture Model

GMM-2.



1. $Y_1 \sim N(\mu_1, \sigma_1^2)$
2. $Y_2 \sim N(\mu_2, \sigma_2^2)$
3. $Y = (1-\Delta)Y_1 + \Delta Y_2$

where $\Delta \sim \text{Bern}(\underbrace{1/2}_{\pi})$



$$P(Y=y) = (1-\pi)\underbrace{N(y;\mu_1,\sigma_1^2)}_{\frac{1}{\sqrt{2\pi}\,\sigma_1}\exp\left(-\frac{(y-\mu_1)^2}{2\sigma_1^2}\right)} + \pi\, N(y;\mu_2,\sigma_2^2)$$

$$\log P(Y=y) = \log\left((1-\pi)\exp(\quad) + \pi\exp(\quad)\right)$$

y_1, δ_1
y_2, δ_2 1st
$\vdots$
y_N, δ_N. 2nd

Y, Δ

$\Delta \in \{0, 1\}$

$\Delta \sim B(\pi)$

y_i, δ_i

$$LL\underbrace{\begin{pmatrix} \pi \\ \mu_1 \\ \sigma_1 \\ \mu_2 \\ \sigma_2 \end{pmatrix}}_{\theta \in \mathbb{R}^5} = \sum_{i=1}^{N} \Big[(1-\delta_i)\left(\log N(y_i; \mu_1, \sigma_1^2) + \log(1-\pi)\right)$$
$$+ \delta_i \left(\log N(y_i; \mu_2, \sigma_2^2) + \log \pi\right) \Big]$$

$$\max_{\pi} \sum_{i=1}^{N} (1-\delta_i)\log(1-\pi) + \delta_i \log \pi$$

M

$$\sum_{i=1}^{N}(1-\delta_i)\frac{-1}{(1-\pi)} + \frac{\sum_{i=1}^{N}\delta_i}{\pi} = 0$$

$$\frac{(N-M)}{1-\pi} = \frac{M}{\pi} \Rightarrow \frac{N-M}{M} = \frac{1}{\pi} - 1$$

$$\Rightarrow \frac{N}{M} = \frac{1}{\pi}$$

$$\Rightarrow \pi = M/N.$$

$$\theta = \begin{bmatrix} \pi \\ \mu_1 \\ \sigma_1 \\ \mu_2 \\ \sigma_2 \end{bmatrix} \qquad \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \mathbf{Y}$$

$$E[\Delta_i \mid \theta, \mathbf{Y}]$$

$$\gamma_i \equiv P(\Delta_i = 1 \mid \theta, y_i)$$

		weight
y_1	1	γ_1
y_1	0	$1-\gamma_1$
y_2	1	γ_2
y_2	0	$1-\gamma_2$.
$\vdots$	$\vdots$	
y_N	1	
y_N	0	

Why EM works:

	$\mathbf{Z}, \mathbf{Z}^m$	Y, Δ
	$\ell(\theta; \mathbf{Z})$ $\ell_0(\theta; \mathbf{Z}, \mathbf{Z}^m)$	
E.	$P(\mathbf{Z}^m \mid \mathbf{Z}, \theta^{old})$	$P(\Delta_i = 1 \mid y_i, \theta^{old}) \quad i = 1 \ldots N$
weighted LL	$Q(\theta', \theta^{old})$	
M.	$\theta^{New} = \arg\max_{\theta'} Q(\theta', \theta^{old})$	

$$P(Z ; \theta^{new}) = \frac{P(Z^m, Z ; \theta^{new})}{P(Z^m | Z, \theta^{new})}$$

$$P(Z^m | Z ; \theta) = \frac{P(Z^m, Z ; \theta)}{P(Z ; \theta)}$$

① log

$$l(Z ; \theta^{new}) = l_0(Z^m, Z ; \theta^{new}) - l_1(\theta^{new})$$

② $E_{P(Z^m | Z, \theta^{old})}$

$$\underbrace{l(Z;\theta^{new})} = \underbrace{Q(\theta^{new},\theta^{old})} - \underbrace{R(\theta^{new},\theta^{old})}$$

$$l(Z,\theta^{old})$$

$$E_{p(z^m|z,\theta^{old})}\left[l_1(\theta^{new})\right]$$

$$\boxed{l(\theta^{new}) - l(\theta^{old}) > 0}$$

$$Q(\theta^{new},\theta^{old}) - R(\theta^{new},\theta^{old}) > \underbrace{Q(\theta^{old},\theta^{old})} - R(\theta^{old},\theta^{old})$$

$$R(\theta^{new}, \theta^{old}) < R(\theta^{old}, \theta^{old})$$

$$\underbrace{\sum_{i=1}^{k} q_i \log p_i} < \underbrace{\sum_{i=1}^{k} q_i \log q_i}$$

$$\equiv \left.\begin{array}{c} \max_{p_i} \sum_{i=1}^{k} q_i \log p_i \\ \sum_{i=1}^{k} p_i = 1 \end{array}\right]$$

$$\mathbb{E}_{\underbrace{p(z^m|\theta^{old})}_{q}} \underbrace{\log p(z^m|\theta^{new})}_{p}$$

$$\sum_{i=1}^{k} q_i \cdot \log p_i$$

$$Obj = \sum_{i=1}^{k} q_i \log p_i - \lambda \left(\sum_{i=1}^{k} p_i - 1 \right) \qquad p_1 \ldots p_k$$

$$\frac{q_i}{p_i} - \lambda(1) = 0 \Rightarrow \frac{q_i}{p_i} = \lambda \Rightarrow q_i = \lambda p_i$$

$$\sum q_i = \lambda \sum p_i$$

$$\sum p_i = 1$$

$$\sum q_i = 1$$

Sampling

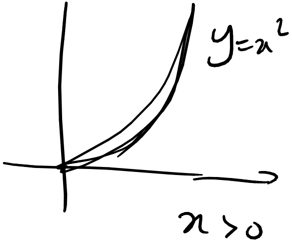
$$P(\beta \mid data) = \frac{P(data \mid \beta) \cdot P(\beta)}{P(data)}$$

$$P(data) = \int_{\beta} P(data \mid \beta) \cdot P(\beta)$$

$$Z \sim P_z \qquad z_1, \ldots, z_N \sim P_z$$

$$E[z] \approx \frac{1}{N} \sum_{i=1}^{N} z_i$$

If we have access to samples from $U[0,1]$



① $Z \sim B(\frac{1}{2})$

$u \sim U[0,1]$

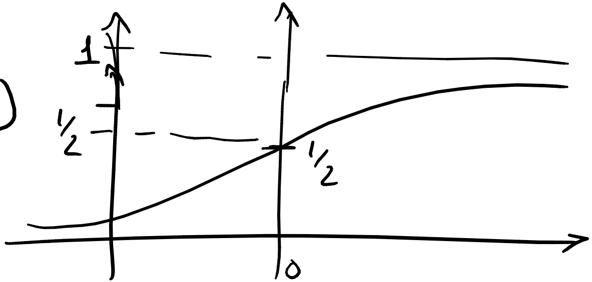
if $u \in [0, 0.5]$

$z = 1$

else

$z = 0$

②



$F(x) = P(X \le x)$

$U \sim P_U$ $\qquad u_1, \ldots u_N$ $\qquad U \in \mathbb{R}^3$

Assume: $P(U_i \mid U_1 \ldots U_{i-1}, U_{i+1}, \ldots U_p)$ Gibbs Sampling.

$$\boxed{\begin{array}{c} P(U_1 \mid U_2 U_3) \\ P(U_2 \mid U_1 U_3) \\ P(U_3 \mid U_1 U_2) \end{array}}$$

Markov Chain
Monte Carlo

technique.

U_1	0	2	3				
U_2	1	4	2				
U_3	-3	-1	-2				
	u_1	u_2					
	$i=1$	$i=2$	$i=3$				

Markov Chain

$$P(z_{next} \mid z_{old}) \begin{bmatrix} 0.3 & .7 \\ .4 & .6 \end{bmatrix}$$

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} .3 & .7 \\ .4 & .6 \end{bmatrix} = \begin{bmatrix} a & b \end{bmatrix}$$