EM Algorithm

Machine Learning Techniques

Karthik Thiagarajan

MLE

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \log \prod_{i=1}^{n} f_X(x_i;\ \boldsymbol{\theta}) \\
&= \sum_{i=1}^{n} \log f_X(x_i;\ \boldsymbol{\theta}) \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right]
\end{aligned}
$$

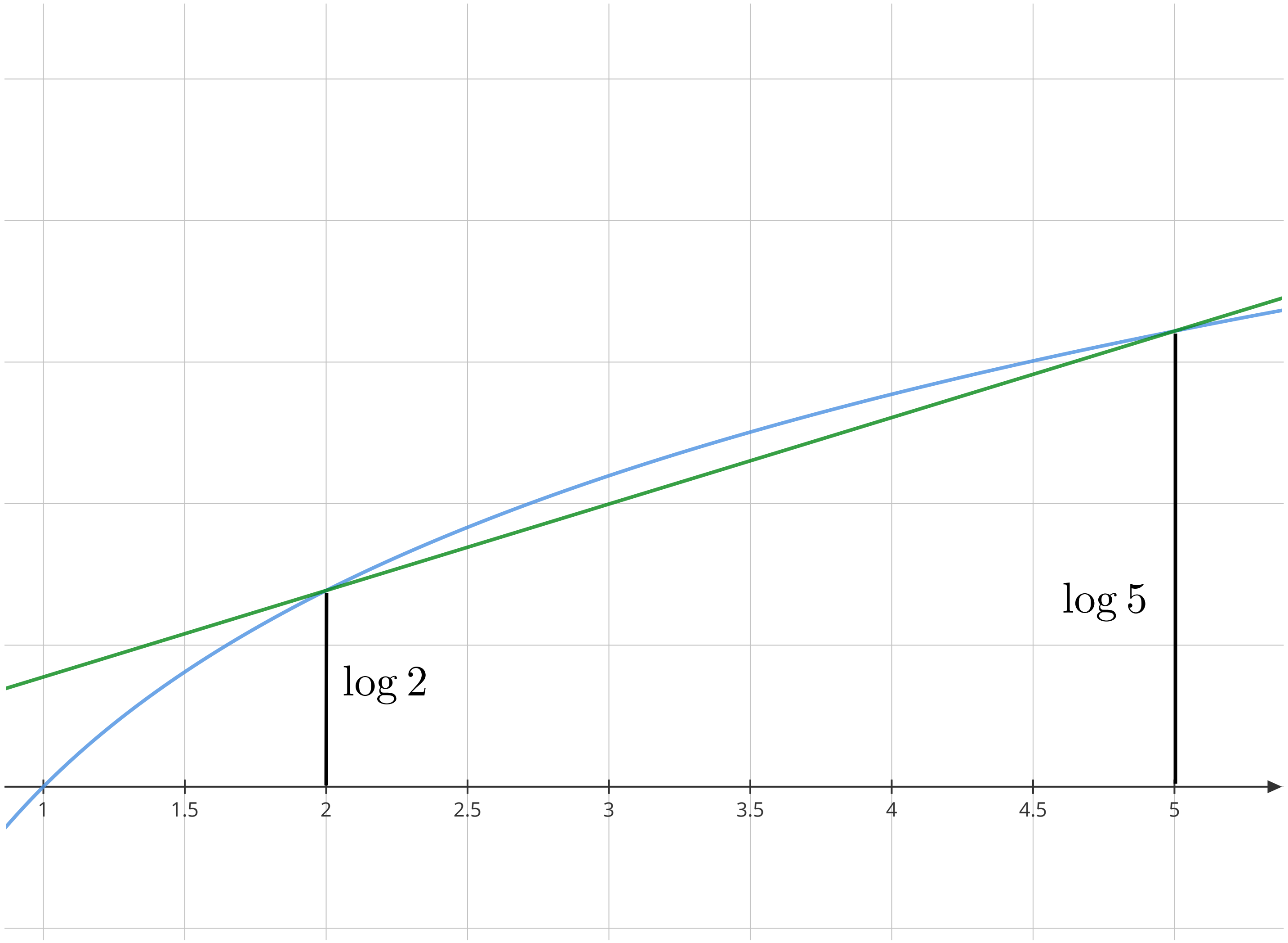
$$\max_{\boldsymbol{\theta}} \quad l(\boldsymbol{\theta}; D)$$

Jensen's Inequality

$$\log\left(\frac{1}{2}\cdot 2+\frac{1}{2}\cdot 5\right)\geq \frac{1}{2}\log 2+\frac{1}{2}\log 5$$

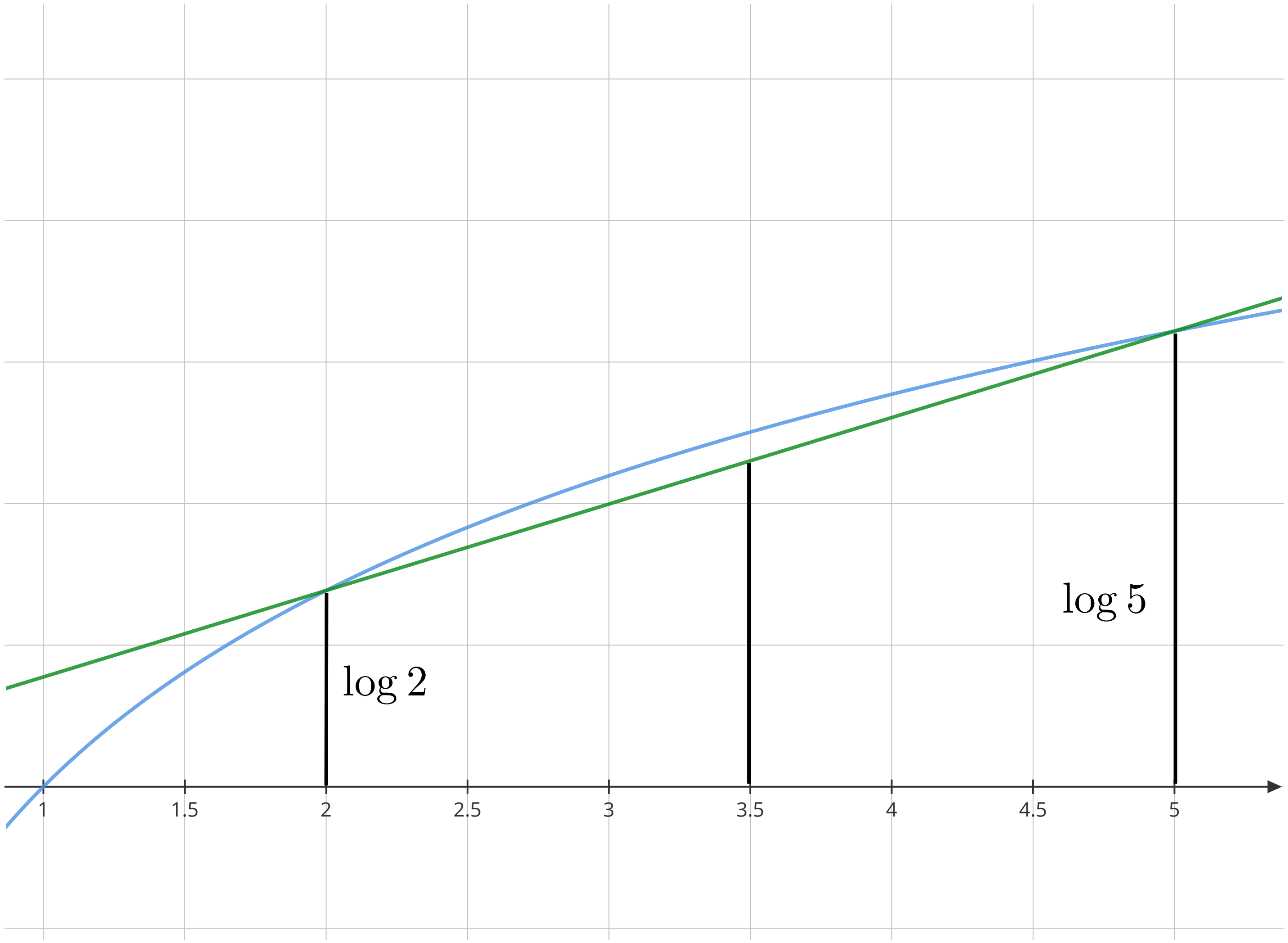
Jensen's Inequality

$$\log\left(\frac{1}{2}\cdot 2 + \frac{1}{2}\cdot 5\right) \geq \frac{1}{2}\log 2 + \frac{1}{2}\log 5$$



Jensen's Inequality

$$\log\left(\frac{1}{2}\cdot 2+\frac{1}{2}\cdot 5\right) \geq \frac{1}{2}\log 2+\frac{1}{2}\log 5$$

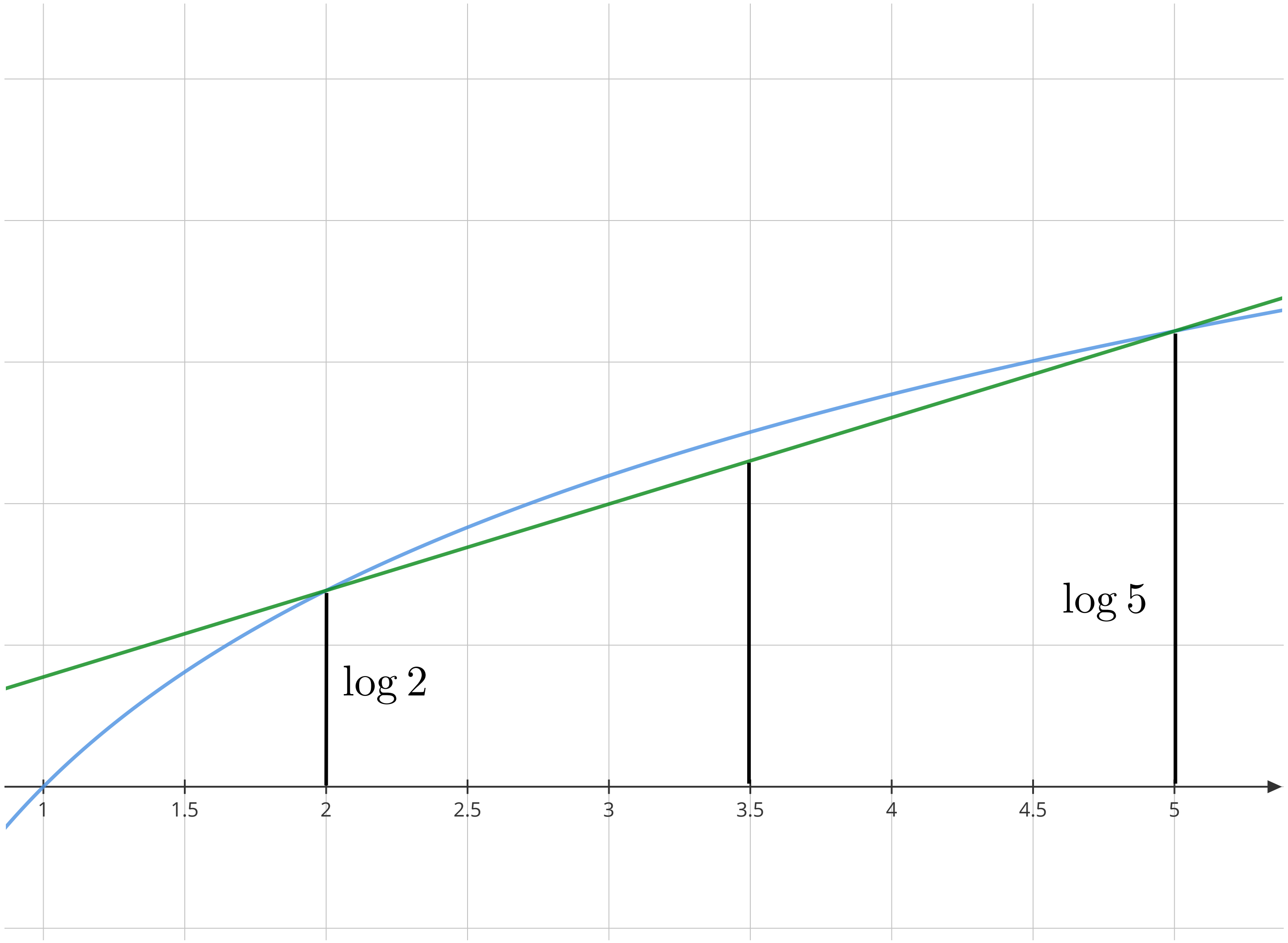


Jensen's Inequality

$$\log\left(\frac{1}{2}\cdot 2+\frac{1}{2}\cdot 5\right) \geq \frac{1}{2}\log 2+\frac{1}{2}\log 5$$

$$\log(\lambda_1 x_1+\lambda_2 x_2) \geq \lambda_1 \log x_1+\lambda_2 \log x_2$$

with $\lambda_1+\lambda_2=1$



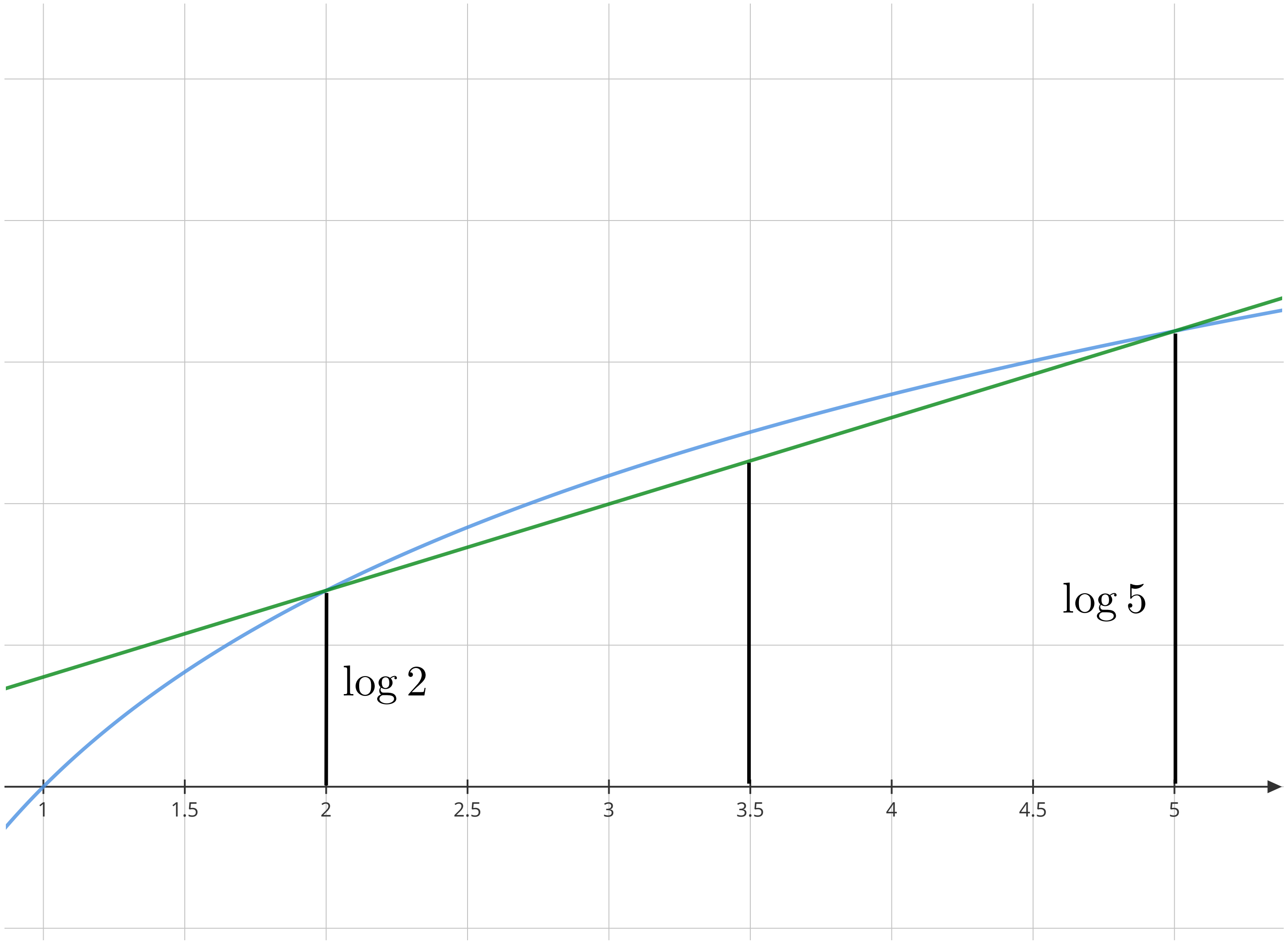
Jensen's Inequality

$$\log\left(\frac{1}{2}\cdot 2+\frac{1}{2}\cdot 5\right) \geq \frac{1}{2}\log 2+\frac{1}{2}\log 5$$

$$\log(\lambda_1 x_1+\lambda_2 x_2) \geq \lambda_1 \log x_1+\lambda_2 \log x_2$$

with $\lambda_1+\lambda_2=1$

Concave → Dome



λ_k^i

$$l(\boldsymbol{\theta};\ D) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right]$$

λ_k^i

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]
\end{aligned}
$$

λ_k^i

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]
\end{aligned}
$$

λ_k^i

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&= \sum_{i=1}^{n} E_{\lambda^i}\left[\log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]\right]
\end{aligned}
$$

λ_k^i

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&= \sum_{i=1}^{n} E_{\lambda^i}\left[\log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]\right]
\end{aligned}
$$

$$\sum_{k=1}^{K} \lambda_k^i = 1$$

$$\lambda_k^i$$

$$\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&= \sum_{i=1}^{n} E_{\lambda^i}\left[\log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]\right]
\end{aligned}$$

$$\sum_{k=1}^{K} \lambda_k^i = 1$$

i

k

$\boldsymbol{\lambda}$	1	2	3	4	5
1	λ_1^1	λ_1^2	λ_1^3	λ_1^4	λ_1^5
2	λ_2^1	λ_2^2	λ_2^3	λ_2^4	λ_2^5
3	λ_3^1	λ_3^2	λ_3^3	λ_3^4	λ_3^5

λ_k^i

$$
\begin{aligned}
l(\boldsymbol{\theta};\ D) &= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)\right] \\
&= \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \lambda_k^i \cdot \frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right] \\
&= \sum_{i=1}^{n} E_{\lambda^i}\left[\log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]\right]
\end{aligned}
$$

$$
\sum_{k=1}^{K} \lambda_k^i = 1
$$

i

k

$\boldsymbol{\lambda}$	1	2	3	4	5
1	λ_1^1	λ_1^2	λ_1^3	λ_1^4	λ_1^5
2	λ_2^1	λ_2^2	λ_2^3	λ_2^4	λ_2^5
3	λ_3^1	λ_3^2	λ_3^3	λ_3^4	λ_3^5

i

k

$\boldsymbol{\lambda}$	1	2	3	4	5
1	0.2	0.5	0.8	0.5	0.3
2	0.7	0.5	0.1	0.3	0.3
3	0.1	0	0.1	0.2	0.4

EM Algorithm

Initialization

$$\boldsymbol{\theta}^{(0)} = \begin{cases} \mu_1^{(0)},\ \cdots, \mu_K^{(0)} \\ \left(\sigma_1^2\right)^{(0)},\ \cdots, \left(\sigma_K^2\right)^{(0)} \\ \pi_1^{(0)},\ \cdots, \pi_K^{(0)} \end{cases}$$

$$\boldsymbol{\theta}^{t} = \begin{cases} \mu_1^{(t)},\ \cdots, \mu_K^{(t)} \\ \left(\sigma_1^2\right)^{(t)},\ \cdots, \left(\sigma_K^2\right)^{(t)} \\ \pi_1^{(t)},\ \cdots, \pi_K^{(t)} \end{cases}$$

EM Algorithm

Initialization

$$\boldsymbol{\theta}^{(0)} = \begin{cases} \mu_1^{(0)},\ \cdots, \mu_K^{(0)} \\ \left(\sigma_1^2\right)^{(0)},\ \cdots, \left(\sigma_K^2\right)^{(0)} \\ \pi_1^{(0)},\ \cdots, \pi_K^{(0)} \end{cases}$$

$$\boldsymbol{\theta}^{t} = \begin{cases} \mu_1^{(t)},\ \cdots, \mu_K^{(t)} \\ \left(\sigma_1^2\right)^{(t)},\ \cdots, \left(\sigma_K^2\right)^{(t)} \\ \pi_1^{(t)},\ \cdots, \pi_K^{(t)} \end{cases}$$

Repeat until convergence:

E-Step:

$$\lambda_k^i = f_{Z|X}(k \mid x_i) = \frac{f_Z(k) \cdot f_{X|Z}(x_i \mid k)}{\sum_{l=1}^{K} f_Z(l) \cdot f_{X|Z}(x_i \mid l)}$$

EM Algorithm

Initialization

$$\boldsymbol{\theta}^{(0)} = \begin{cases} \mu_1^{(0)},\ \cdots, \mu_K^{(0)} \\ \left(\sigma_1^2\right)^{(0)},\ \cdots, \left(\sigma_K^2\right)^{(0)} \\ \pi_1^{(0)},\ \cdots, \pi_K^{(0)} \end{cases}$$

$$\boldsymbol{\theta}^{t} = \begin{cases} \mu_1^{(t)},\ \cdots, \mu_K^{(t)} \\ \left(\sigma_1^2\right)^{(t)},\ \cdots, \left(\sigma_K^2\right)^{(t)} \\ \pi_1^{(t)},\ \cdots, \pi_K^{(t)} \end{cases}$$

Repeat until convergence:

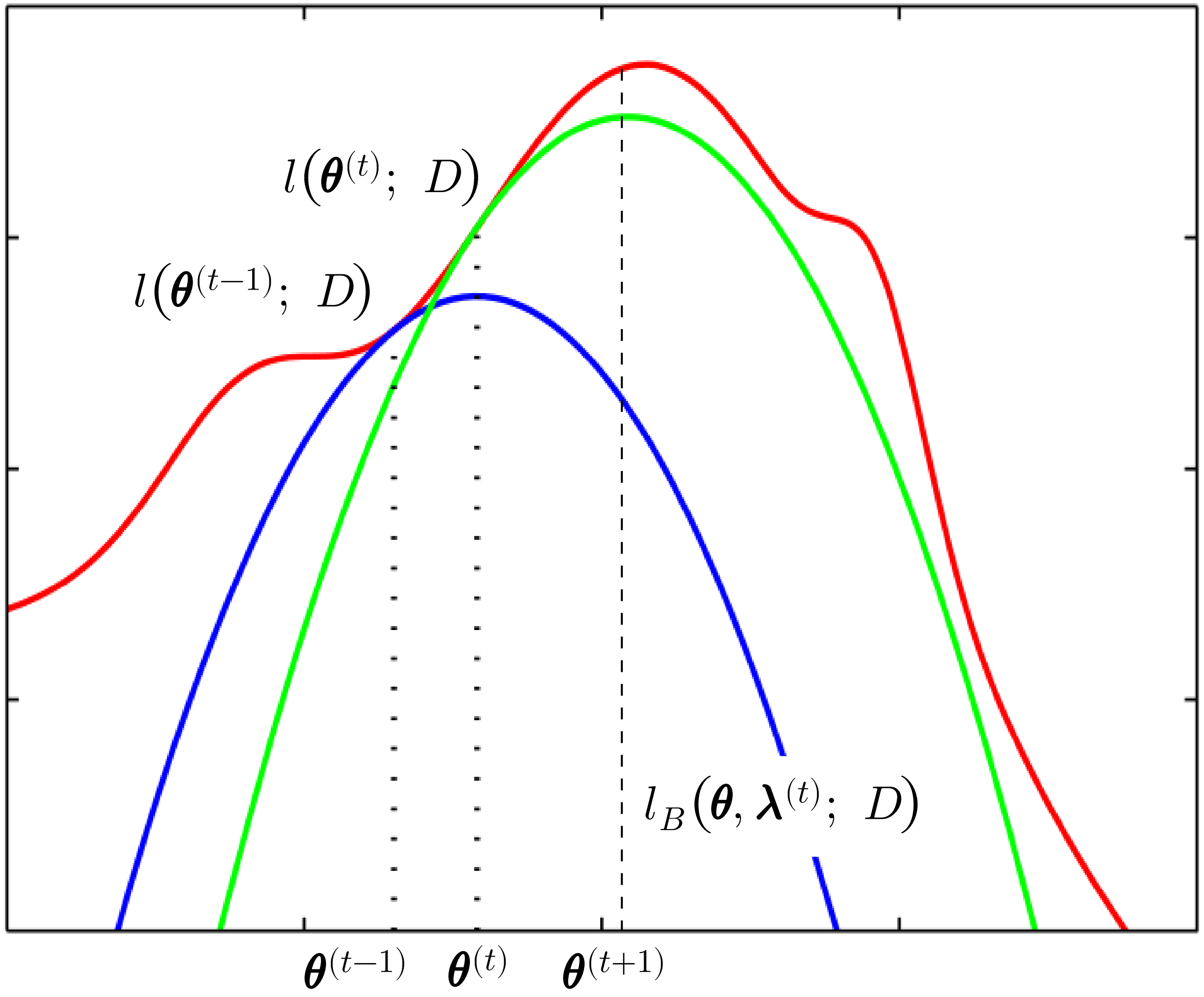
E-Step:

$$\lambda_k^i = f_{Z|X}(k \mid x_i) = \frac{f_Z(k) \cdot f_{X|Z}(x_i \mid k)}{\sum_{l=1}^{K} f_Z(l) \cdot f_{X|Z}(x_i \mid l)}$$

M-Step:

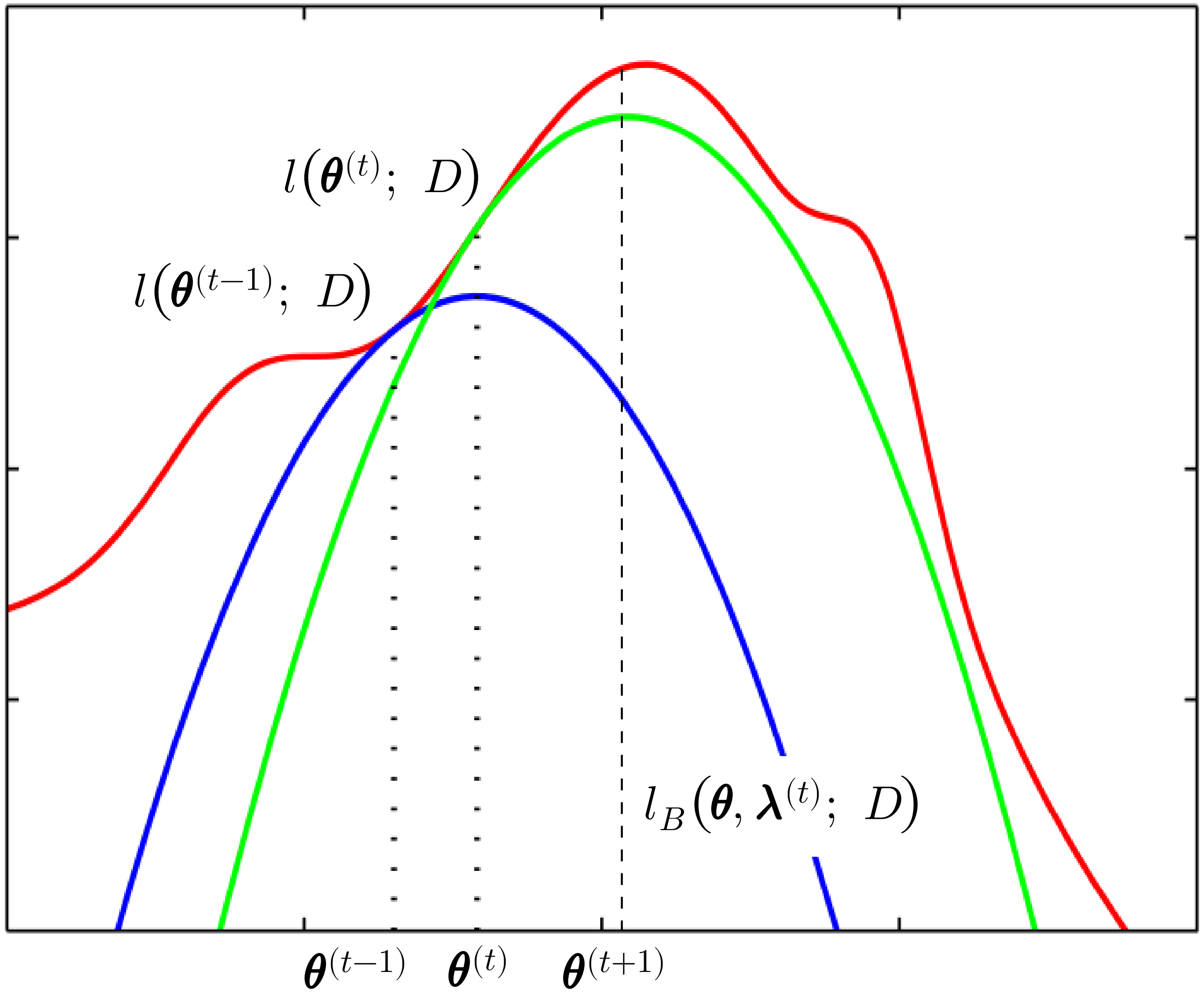
$$\mu_k^{(t)} = \frac{\sum_{i=1}^{n} \lambda_k^i x_i}{\sum_{i=1}^{n} \lambda_k^i} \qquad \left(\sigma_k^2\right)^{(t)} = \frac{\sum_{i=1}^{n} \lambda_k^i \left(x_i - \mu_k^{(t)}\right)^2}{\sum_{i=1}^{n} \lambda_k^i} \qquad \pi_k^{(t)} = \frac{\sum_{i=1}^{n} \lambda_k^i}{n}$$

EM Algorithm



Chapter 9.4,Bishop, Pg 423

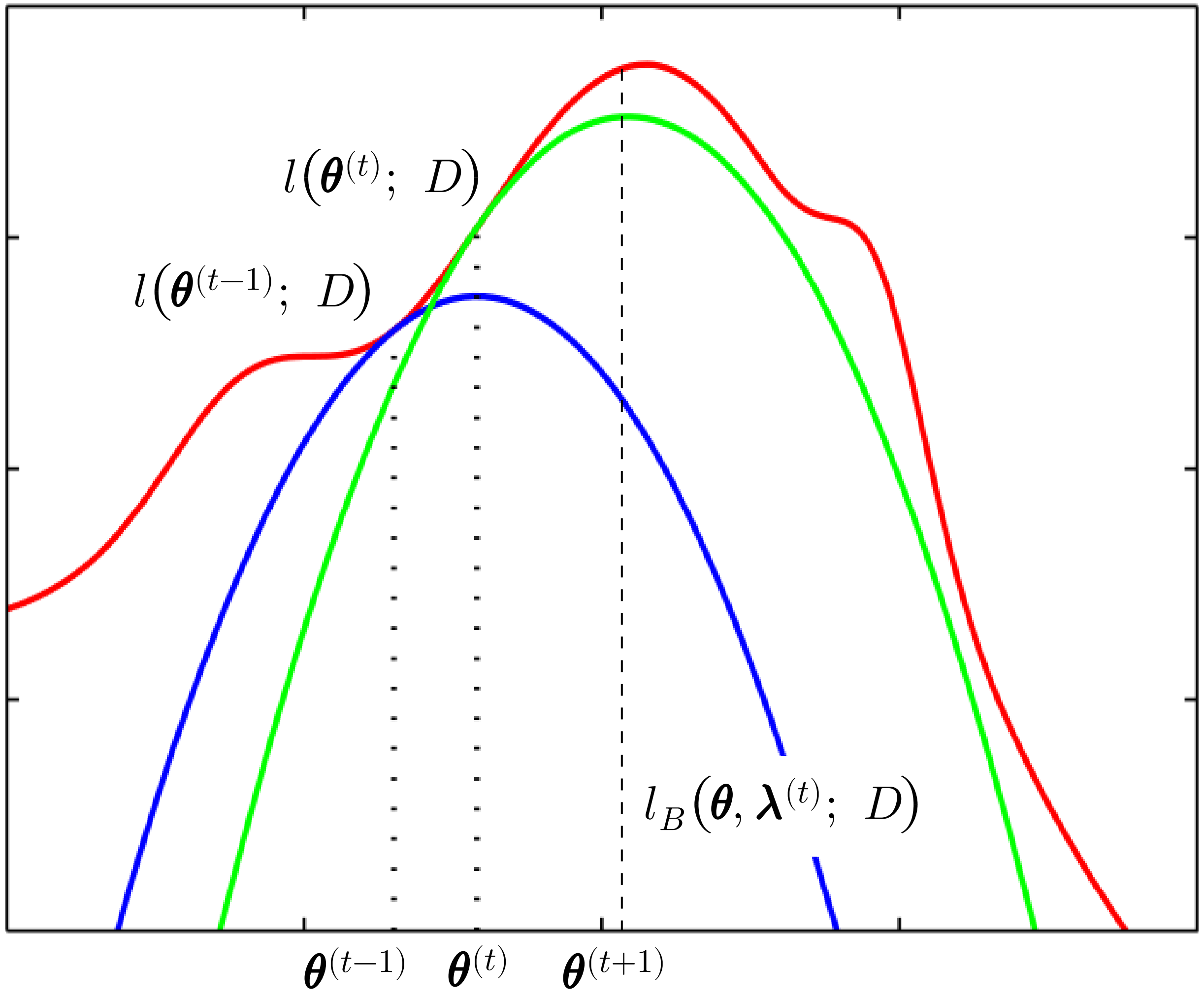
EM Algorithm



Chapter 9.4,Bishop, Pg 423

$$l(\boldsymbol{\theta}^{(t)};\ D) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k^{(t)} \cdot \mathcal{N}\left(x_i;\ \mu_k^{(t)}, \left(\sigma_k^2\right)^{(t)}\right)\right]$$

EM Algorithm

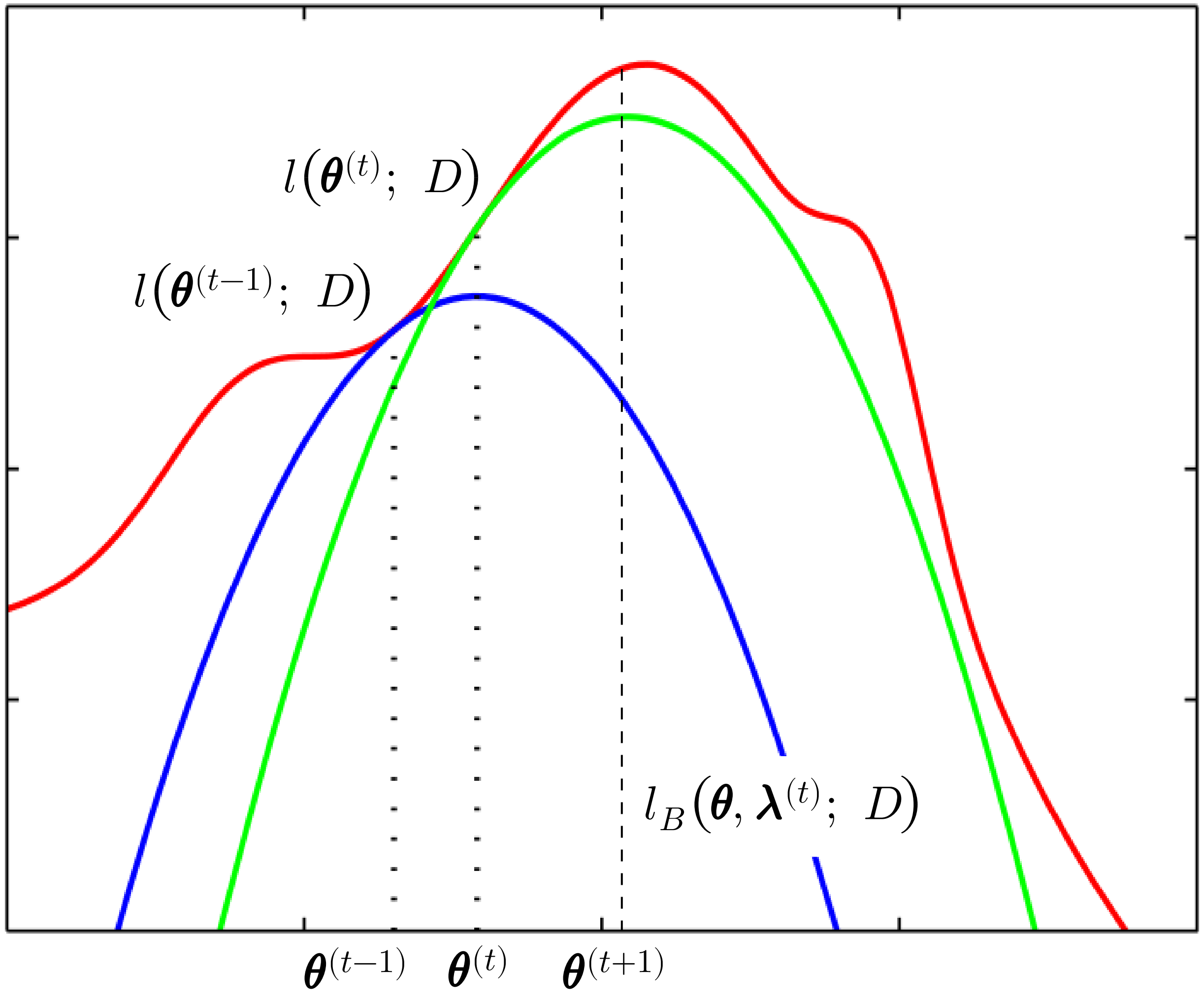


Chapter 9.4,Bishop, Pg 423

$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k^{(t)} \cdot \mathcal{N}\left(x_i;\ \mu_k^{(t)}, \left(\sigma_k^2\right)^{(t)}\right)\right]$$

$$l_B(\boldsymbol{\theta}, \boldsymbol{\lambda};\ D) = \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]$$

EM Algorithm



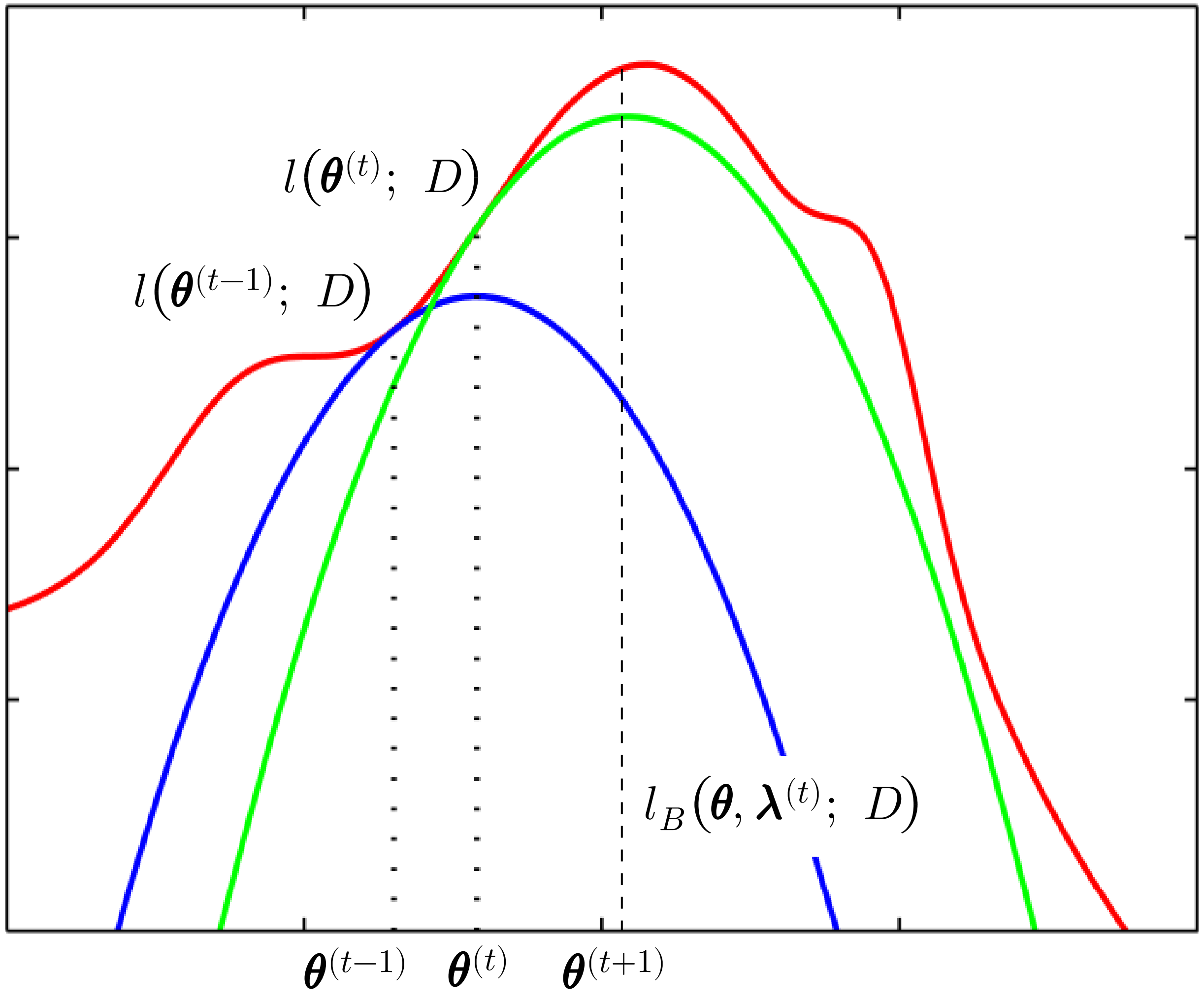
Chapter 9.4,Bishop, Pg 423

$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k^{(t)} \cdot \mathcal{N}\left(x_i;\ \mu_k^{(t)}, \left(\sigma_k^2\right)^{(t)}\right)\right]$$

$$l_B(\boldsymbol{\theta}, \boldsymbol{\lambda};\ D) = \sum_{i=1}^{n}\sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]$$

E-step: $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}}\ l_B\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda};\ D\right)$

EM Algorithm



Chapter 9.4,Bishop, Pg 423

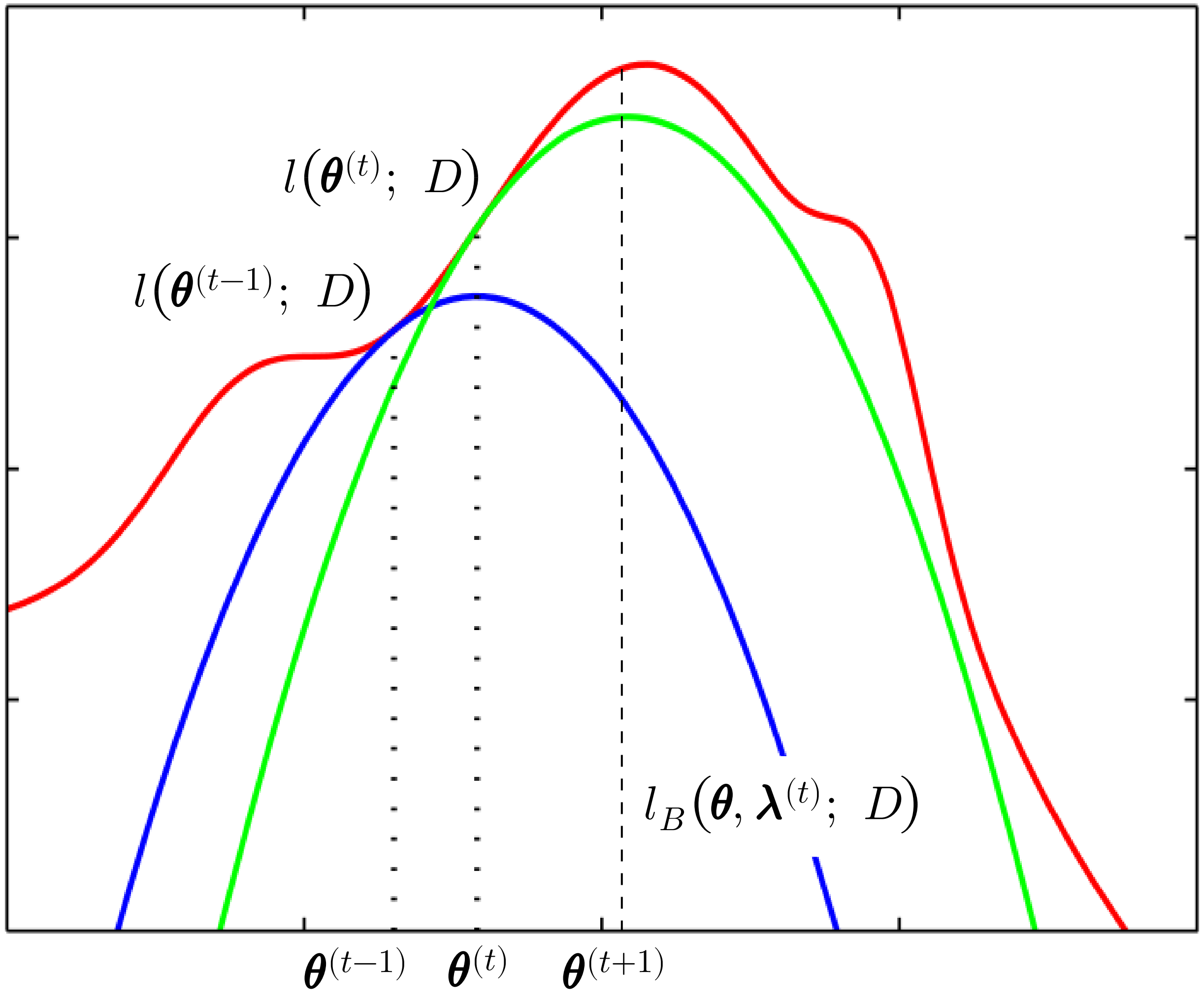
$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k^{(t)} \cdot \mathcal{N}\left(x_i;\ \mu_k^{(t)}, \left(\sigma_k^2\right)^{(t)}\right)\right]$$

$$l_B(\boldsymbol{\theta}, \boldsymbol{\lambda};\ D) = \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]$$

E-step: $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}}\ l_B\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda};\ D\right)$

$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = l_B\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda}^{(t+1)};\ D\right)$$

EM Algorithm



Chapter 9.4,Bishop, Pg 423

$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = \sum_{i=1}^{n} \log\left[\sum_{k=1}^{K} \pi_k^{(t)} \cdot \mathcal{N}\left(x_i;\ \mu_k^{(t)}, \left(\sigma_k^2\right)^{(t)}\right)\right]$$

$$l_B(\boldsymbol{\theta}, \boldsymbol{\lambda};\ D) = \sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k^i \cdot \log\left[\frac{\pi_k \cdot \mathcal{N}\left(x_i;\ \mu_k, \sigma_k^2\right)}{\lambda_k^i}\right]$$

E-step: $\boldsymbol{\lambda}^{(t+1)} = \arg\max_{\boldsymbol{\lambda}}\ l_B\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda};\ D\right)$

$$l\left(\boldsymbol{\theta}^{(t)};\ D\right) = l_B\left(\boldsymbol{\theta}^{(t)}, \boldsymbol{\lambda}^{(t+1)};\ D\right)$$

M-step: $\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}}\ l_B\left(\boldsymbol{\theta}, \boldsymbol{\lambda}^{(t+1)};\ D\right)$

EM Algorithm

K-Means

E-step: $\lambda_k^i = f_{Z|X}(k \mid x_i)$

EM Algorithm

E-step: $\lambda_k^i = f_{Z|X}(k \mid x_i)$

K-Means

Reassignment-step: $\lambda_k^i = \begin{cases} 1 & x_i \text{ is closest to } \mu_k \\ 0 & \text{otherwise} \end{cases}$

EM Algorithm

E-step: $\lambda_k^i = f_{Z|X}(k \mid x_i)$

M-step:

$$\mu_k = \frac{\sum_{i=1}^{n} \lambda_k^i x_i}{\sum_{i=1}^{n} \lambda_k^i}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{n} \lambda_k^i (x_i - \mu_k)^2}{\sum_{i=1}^{n} \lambda_k^i} \qquad \pi_k = \frac{\sum_{i=1}^{n} \lambda_k^i}{n}$$

K-Means

Reassignment-step: $\lambda_k^i = \begin{cases} 1 & x_i \text{ is closest to } \mu_k \\ 0 & \text{otherwise} \end{cases}$

EM Algorithm

E-step: $\lambda_k^i = f_{Z|X}(k \mid x_i)$

M-step:

$$\mu_k = \frac{\sum_{i=1}^{n} \lambda_k^i x_i}{\sum_{i=1}^{n} \lambda_k^i}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{n} \lambda_k^i (x_i - \mu_k)^2}{\sum_{i=1}^{n} \lambda_k^i} \qquad \pi_k = \frac{\sum_{i=1}^{n} \lambda_k^i}{n}$$

K-Means

Reassignment-step: $\lambda_k^i = \begin{cases} 1 & x_i \text{ is closest to } \mu_k \\ 0 & \text{otherwise} \end{cases}$

Compute mean: $\mu_k = \dfrac{\sum_{i=1}^{n} \lambda_k^i \cdot x_i}{\sum_{i=1}^{n} \lambda_k^i}$