COL774
Machine Learning
Oct 27, 2021

Last Class:- EM expectation maximization

$\{x^{(i)}\}_{i=1}^{m}$ $\quad P(x^{(i)}, z^{(i)}; \theta)$ ↳ hidden

$LL(\theta) = \sum_{i=1}^{m} \log P(x^{(i)}; \theta)$ ↳ intractable (using gradient descent)

$\Downarrow$

$\sum_{i=1}^{m} \log \sum_{z^{(i)}} P(x^{(i)}, z^{(i)}; \theta)$

Jensen's inequality: for f concave

$$f(E[X]) \geq E[f(x)]$$

$\Rightarrow$

$$LL(\theta) = \sum_{i=1}^{m} \log \sum_{z^{(i)}} \left[\frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \cdot \right] Q_i(z^{(i)})$$

$$= \sum_{i=1}^{m} \log E_{Q_i}\left[\frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right] \to C$$

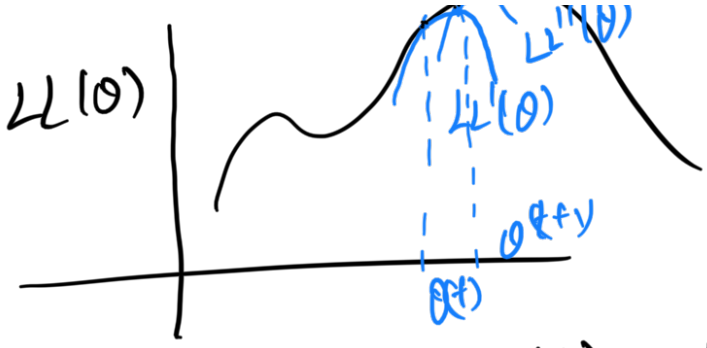
↳ concave

E-step

$$\geq \sum_{i=1}^{m} E_{Q_i}\left[\log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right] \Rightarrow C$$

$$= \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} = LL'(\theta)$$

$LL(\theta) \geq LL'(\theta)$ → Lower bound



$\max_\theta LL''(\theta) \geq \max_\theta LL'(\theta)$

- - -

When does $LL(\theta) = LL'(\theta)$

If $\frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} = C \quad \forall i \quad -①$

$\Rightarrow P(x^{(i)}, z^{(i)}; \theta) = C\, Q_i(z^{(i)}) \quad \forall i$

$\Rightarrow \sum_{z^{(i)}} P(x^{(i)}, z^{(i)}; \theta) = C \sum_{z^{(i)}} Q_i(z^{(i)})$ (sum over $z^{(i)}$'s)

$\Rightarrow P(x^{(i)}; \theta) = C \cdot 1$

$\Rightarrow Q_i(z^{(i)}) = \frac{P(x^{(i)}, z^{(i)}; \theta)}{C} = \frac{P(x^{(i)}, z^{(i)}; \theta)}{P(x^{(i)}; \theta)}$

$= P(z^{(i)} | x^{(i)}; \theta)$

$\Rightarrow$ If we "fill in" missing values s.t

$Q_i(z^{(i)}) = P(z^{(i)} | x^{(i)}; \theta)$

then the lower bound will touch the function at the "correct" value of θ parameters

M-step: E-step:- construct $LL'(\theta)$

using $Q_i(z^{(i)}) = P(z^{(i)} | x^{(i)}; \theta)$

M. step:- $\arg\max_\theta LL'(\theta)$

$= \sum_i \log(\quad)$ for $P(x^{(i)}, z^{(i)}; \theta)$?

$$LL'(\theta) = \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \left[\log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right]$$

If not very different to optimize.

$$\arg\max_\theta LL'(\theta) = \arg\max_\theta \left[\sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log P(x^{(i)}, z^{(i)}; \theta) + \cancel{\text{Entropy}} \right]$$

does not depend on θ

Entropy of Q_i: $\sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log Q_i(z^{(i)})$

$$\equiv \arg\max_\theta \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log P(x^{(i)}, z^{(i)}; \theta)$$

(Contrast with exp GMM)

Repeat

EM Algorithm:-

$\{x^{(i)}\}_{i=1}^{m}$ $x^{(i)} \in \mathbb{R}^n$

$P(x^{(i)}, z^{(i)}; \theta)$

~~Probability~~

$t \leftarrow 0;$

$\theta^{(t)} \leftarrow$ init();

do {

E-Step :- $\forall i$:- $Q_i^{t}(z^{(i)}) \leftarrow P(z^{(i)} | x^{(i)}; \theta^{(t)})$

M-Step :- $\theta^{(t+1)} \leftarrow \arg\max_\theta LL^{t}(\theta);$

$\equiv \arg\max_\theta \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i^{t}(z^{(i)}) \log \frac{P(x^{(i)}, z^{(i)}; \theta)}{\cancel{Q_i(z^{(i)})}}$

$t \leftarrow t+1;$

} while (! converged)

Convergence:- $\|\theta^{(t+1)} - \theta^{(t)}\| \leq \epsilon_0$

↳ $\nabla_\theta LL^t(\theta)\big|_{\theta^{(t)}} \simeq 0$

$\Rightarrow \theta^{(t+1)} \simeq \theta^{(t)}$

Note:- ① Converges to local optima of $LL(\theta)$ ✓

② $\arg\max_\theta LL(\theta)$ ↳ EM can be seen as optimizing $\hat{LL}(\theta, \theta_1)$

$\equiv \arg\max_{\theta, \theta_1} \hat{LL}(\theta, \theta_1)$ using block coordinate descent

$$\hat{LL}(\theta, \theta_1) = \sum_{i=1}^{M} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \left[\frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right]$$

Recall EM:- Algorithm

- M-step correspond to:- $\max_\theta \hat{LL}(\theta, \theta_1)$ (For a given θ_1)
- E-step correspond to:- $\max_{\theta_1} \hat{LL}(\theta, \theta_1)$ (For a given θ)

Recall:- $LL(\theta) \geq \hat{LL}(\theta, \theta_1) \quad \forall \theta_1$

(By construction / Jensen's inequality)

$\Rightarrow$ EM is nothing but Block coordinate

Result $LL(\theta, \theta^{(t)})$

EM converges to local optima of $LL(\theta)$

$$LL^t(\theta^{(t)}) \geq \cdots \geq LL^2(\theta^{(2)}) \geq LL^1(\theta^{(1)}) \geq LL^0(\theta^{(0)}) \quad -(i)$$

$$\forall t \quad LL(\theta^{(t)}) = LL^t(\theta^{(t)}) \quad \forall t \quad -(ii)$$

Suppose EM has converged.

$$\theta^{(t+1)} \simeq \theta^{(t)}$$

$$\Rightarrow \nabla_\theta LL^t(\theta)\big|_{\theta^{(t)}} = 0 \quad -(iii)$$

(By (i))

$(\because \nabla_\theta LL^t(\theta)\big|_{\theta^{(t+1)}} = 0$ By const)

Claim:- $\nabla_\theta LL(\theta)\big|_{\theta^{(t)}} = 0 \Rightarrow$ Local optima

Assume o.w.:- $\nabla_\theta LL(\theta)\big|_{\theta^{(t)}} > 0$

$$\Delta\theta = \eta \cdot \nabla_\theta LL(\theta)\big|_{\theta^{(t)}}$$

$$\frac{LL(\theta^{(t)} + \Delta\theta) - LL(\theta^{(t)})}{\Delta\theta} > 0$$

$$\frac{LL(\theta^{(t)} - \Delta\theta) - LL(\theta^{(t)})}{\Delta\theta} < 0 \quad -(iv)$$

$$\frac{LL^t(\theta^{(t)} - \Delta\theta) - LL^t(\theta^{(t)})}{\Delta\theta} = 0 \quad (v)$$

< 0

& ④ - ⑤

$$LL(\theta^{(t)} - \Delta\theta) - LL(\theta^{(t)}) - \left[LL^t(\theta^{(t)} - \Delta\theta) - LL^t(\theta^{(t)})\right] < 0$$

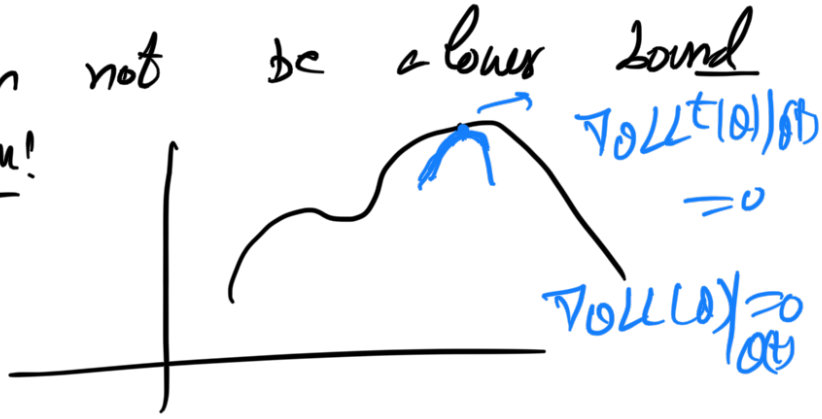
$$\Delta\theta$$

$$\Rightarrow \quad \frac{LL(\theta^{(t)} - \Delta\theta) - LL^t(\theta^{(t)} - \Delta\theta)}{\Delta\theta} < 0$$

$$\Rightarrow \quad LL(\theta^{(t)} - \Delta\theta) - LL^{(t)}(\theta^{(t)} - \Delta\theta) < 0$$

$\Rightarrow$ $LL^{(t)}$ can not be a lower bound

$\Rightarrow$ contradiction!



$\Rightarrow$ EM algorithm:-

PCA:- Principal Component Analysis

(Next class)