Coursera — Bayesian Methods for ML

EM Algo

Latent Variable $(t_i) \to (x_i)$

$$p(x_i|\theta) = \sum_c p(x_i|t_i=c,\theta)\, p(t_i=c|\theta)$$

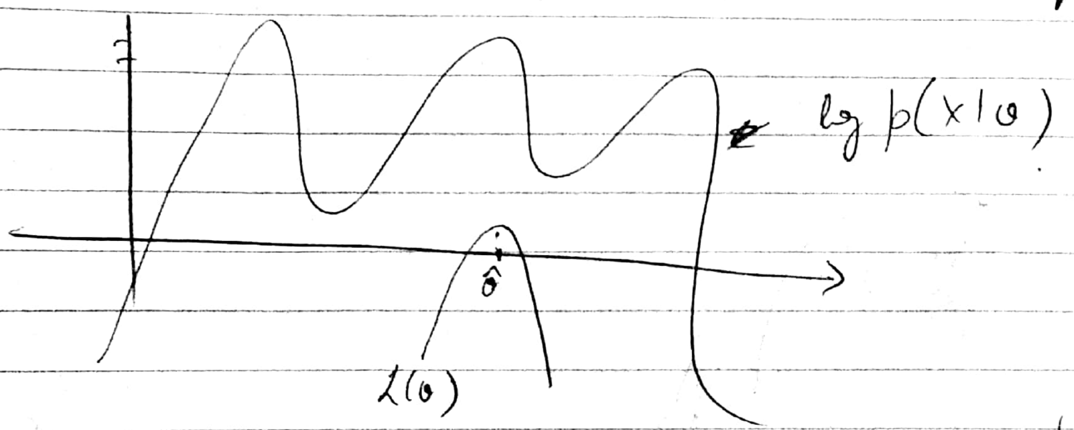
Goal $\max_\theta p(X|\theta) = \max_\theta \prod_i p(x_i|\theta)$

$\Rightarrow$

$$\max_\theta \log p(X|\theta) = \max_\theta \left(\sum \log p(x_i|\theta)\right)$$

$$\sum \log p(x_i|\theta) = \sum \log \sum p(x_i, t_i=c|\theta) + k \geq \mathcal{L}(\theta)$$

(Jensen's Inequality)



$\therefore \mathcal{L}(\theta)$ may get max at arbitrary value! (Variational distri)

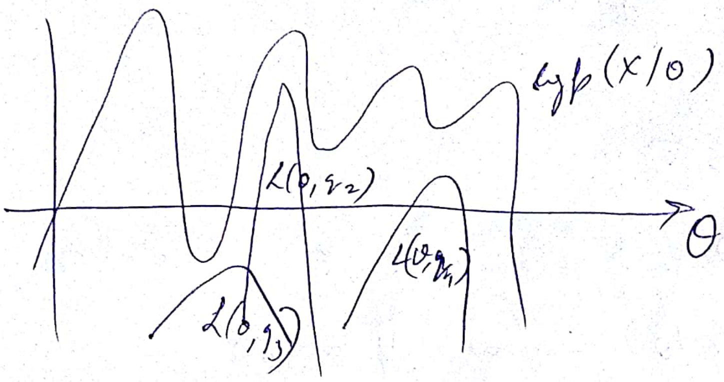
Better $\log p(X|\theta) = \sum_i \log \sum_c \frac{q(t_i=c)}{q(t_i=c)} p(x_i, t_i=c|\theta)$

> Note Jensen's inequality:-
>
> $$\log\left(\sum_c \alpha_c v_c\right) \geq \sum_c \alpha_c \log(v_c)$$

$$\therefore \log p(X|\theta) \geq \sum_i \sum_c q(t_i=c) \log\left(\frac{p(x_i, t_i=c|\theta)}{q(t_i=c)}\right)$$

$\log p(X/\theta) \geq \mathcal{L}(\theta, q)$ for any q. ②

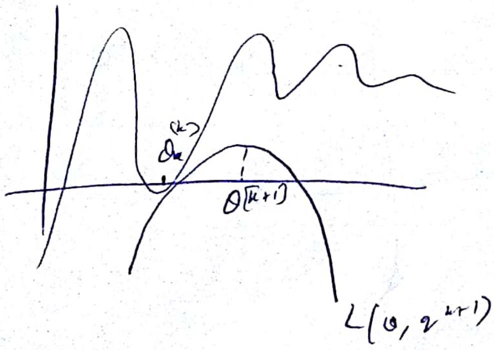
$\Rightarrow \mathcal{L} \equiv$ Variational lower bound.



$q_2 \equiv$ best. How to find q_i?

- start with any point $\theta^{[k]}$

  find $q^{[k+1]} = \arg\max_{q} \mathcal{L}(\theta^{[k]}, q)$



$\Rightarrow$ EM Algo

(1) $\log p(x/\theta) \geq \mathcal{L}(\theta, q)$ for any q.

(2) E-Step $q^{k+1} = \arg\max_{q} \mathcal{L}(\theta^{k}, q)$

(3) M-step : $\theta^{k+1} = \arg\max_{\theta} \mathcal{L}(\theta, q^{k+1})$

E-step $\max_q \mathcal{L}(\theta, q)$

$$Gap = \log p(X|\theta) - \mathcal{L}(\theta, q)$$

$$= \sum_i \log p(x_i|\theta) - \sum_i \sum_c q(t_i=c) \log\left(\frac{p(x_i, t_i=c|\theta)}{q(t_i=c)}\right)$$

$$= \sum_i \left[\log p(x_i|\theta) \underbrace{\sum_c q(t_i=c)}_{1} - \sum_c q(t_i=c) \log\left(\frac{p(x_i, t_i=c|\theta)}{q(t_i=c)}\right)\right]$$

$$= \sum_i \sum_c q(t_i=c)\left(\log p(x_i|\theta) - \log\left(\frac{p(x_i, t_i=c|\theta)}{q(t_i=c)}\right)\right)$$

$$= \sum_i \sum_c q(t_i=c) \log\left(\frac{p(x_i|\theta)\, q(t_i=c)}{p(x_i, t_i=c|\theta)}\right)$$

$$= \sum_i \underbrace{\sum_c q(t_i=c) \log\left(\frac{q(t_i=c)}{p(t_i=c|x_i,\theta)}\right)}_{KL\left(q(t_i)\,||\,p(t_i|x_i,\theta)\right)}$$

$$= \sum_i KL\left(q(t_i)\,||\,p(t_i|x_i,\theta)\right) \geq 0$$

& $KL\left(q(t_i)\,||\,p(t_i|x_i,\theta)\right) = 0$ when $q(t_i) = p(t_i|x_i,\theta)$

$$\therefore \arg\min_q (Gap) = q(t_i) = p(t_i|x_i,\theta)$$

$$\boxed{q^{(k+1)} = p(t_i|x_i,\theta^{(k)})}$$

5

M-step

$$\theta^{k+1} = \underset{\theta}{\arg\max}\ \mathcal{L}(\theta, q^{k+1})$$

$$= \underset{\theta}{\arg\max}\left[\sum_i \sum_c q(t_i=c) \log p(x_i, t_i=c|\theta) - \underbrace{\sum_i \sum_c q(t_i=c) \log q(t_i=c)}_{\text{constt.}}\right]$$

$$= \underset{\theta}{\arg\max} \sum_i \sum_c q(t_i=c) \log p(x_i, t_i=c|\theta)$$

$$= \underset{\theta}{\arg\max}\left(\underbrace{\mathbb{E}_q \log p(X, T|\theta)}_{\text{usually concave fn.}} + \text{constt}\right)$$

(taken for ease of optimization)

Note $\log p(X|\theta^{k+1}) \geq \mathcal{L}(\theta^{k+1}, q^{k+1}) \geq \mathcal{L}(\theta^k, q^{k+1})$

$\because \theta^{k+1} = \underset{\theta}{\arg\max}\ \mathcal{L}$

Also, $\mathcal{L}(\theta^k, q^{k+1}) \leq \log p(X|\theta^k)$

$\mathcal{L}(\theta^k, q^{k+1}) = \max_q \mathcal{L}(\theta^k, q)$

$\Rightarrow \mathcal{L}(\theta^k, q^{k+1}) = \log p(X|\theta^k)$ (at best)

$$\Rightarrow \boxed{\log p(X|\theta^{k+1}) \geq \log p(X|\theta^k) \quad \text{(always)}}$$

$\Rightarrow$ guaranteed to converge.

Konark.

Properties

- Latent Variable Method.
- Handles Missing data.
- Simple instead of lard of fimigolia pur.
- Guaranteed convergence
- Helps with complicated parameter constraints
- Many extensions.

Cons

- Only local Maxima.