EM & Factor Analysis

- Introduce general EM Algorithm
- Gaussian mixture as EM
- Factor Analysis

Recap

We examined GMMs (light sources)

Given $x^{(1)} \dots x^{(n)} \in \mathbb{R}^d$ and $k > 0$

Do: find $P(z^{(i)} = j)$ for $i = 1 \dots n$, $j = 1 \dots k$

We called $z^{(i)}$ a latent variable (not observed directly)

Today: Derive GMM Algo in more general way

Keep: E-step, M-step

Guess $z^{(i)}$ ← E-step

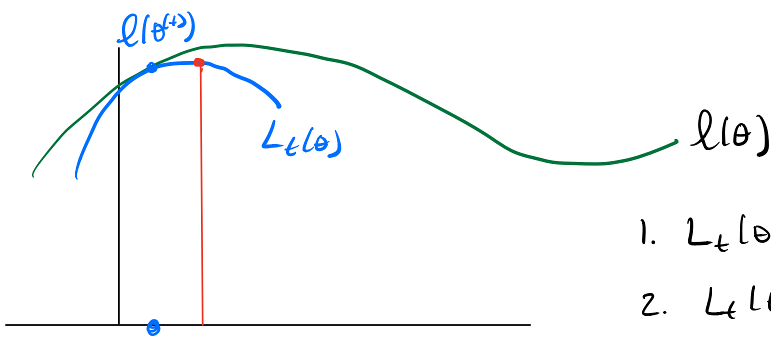
M-step → fit parameters of model

Latent variable model

$$\ell(\theta) = \sum_{i=1}^{n} \log P(\underbrace{x^{(i)}}_{\text{DATA}}; \underbrace{\theta}_{\text{PARAMETERS}})$$

$$= \sum_{i=1}^{n} \sum_{z} \log P(x^{(i)} \mid \underbrace{z^{(i)} = z}_{\text{latent variable}}; \theta) P(z^{(i)} = z; \theta)$$

Picture of our Algorithm (cf w/ GMM)



$\theta^{(t)}$ $\theta^{(t+1)} := \operatorname{Argmax} L_t(\theta)$

Our Estimate at t

1. $L_t(\theta) \leq \ell(\theta)$ (lowerbound)
2. $L_t(\theta^{(t)}) = \ell(\theta^{(t)})$ (tight)

Hope: $L_t(\theta)$ easier to optimize than $\ell(\theta)$)

(E-step) find $L_t(\theta)$ given $\theta^{(t)}$

(M-step) $\theta^{(t+1)} = \operatorname{Argmax} L_t(\theta)$

Why does this Abstract GMM?

Next: How do we find $L_t(\theta)$ given $\theta^{(t)}$

IDEA WE go term-by term $\log P(x^{(i)};\theta)$ Single term

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

let $Q^{(i)}(z)$ s.t. $\sum_z Q^{(i)}(z) = 1$, $Q^{(i)}(z) \geq 0$ $(*)$

Symbol pushing! Just def of $\mathbb{E}$

$$= \log \sum_z \frac{Q^{(i)}(z)}{Q^{(i)}(z)} P(x^{(i)}, z^{(i)})$$

$$= \log \mathbb{E}\left[\frac{P(x^{(i)}, z^{(i)})}{Q^{(i)}(z)}\right]$$

RECALL JENSEN, WE CAN SWITCH $\log(\mathbb{E}[x]) \geq \mathbb{E}[\log(x)]$

(SIDEBAR, next page)

$$\geq \mathbb{E}_z\left[\log \frac{P(x^{(i)}, z^{(i)})}{Q^{(i)}(z)}\right]$$

$$= \sum_z Q^{(i)}(z) \log \frac{P(x^{(i)}, z^{(i)})}{Q^{(i)}(z)}$$

for ANY $Q^{(i)}$ ABOVE this holds, AND for EACH term

So Pick Any $Q^{(i)} \to$ gives AN $L_t(\theta)$!

CAll this $\text{ELBO}(x, Q;\theta) = \sum_z Q(z) \log \frac{P(x, z;\theta)}{Q(z)}$

SHOWN $P(x^{(i)};\theta) \geq \text{ELBO}(x, Q^{(i)};\theta)$

I.e. Property 1

Property 2

WE PICK A SPECIFIC $Q^{(i)}$ depending on $x^{(i)}$ & $\theta^{(t)}$

So that $\ell(\theta^{(t)}) = L_t(\theta^{(t)})$

Goal: Pick $Q^{(i)}$ s.t.

$$\log \sum_z P(x^{(i)}, z^{(i)}) = \sum_z Q^{(i)}(z) \log \frac{P(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q^{(i)}(z)}$$

SET $Q^{(i)}(z) = P(z \mid x^{(i)}; \theta^{(t)})$

$\Rightarrow$

DOESN'T DEPEND ON Z

$$\frac{P(x^{(i)}, z^{(i)})}{Q^{(i)}(z)} = \frac{\cancel{P(z^{(i)} \mid x^{(i)}; \theta^{(t)})}\ \underbrace{P(x^{(i)}; \theta^{(t)})}_{C}}{\cancel{P(z^{(i)} \mid x^{(i)}; \theta^{(t)})}}$$

Hence $\sum_z Q^{(i)}(z) \cdot \log C = \log C$ (RHS/ELBO)

(LHS) $\log \sum_z P(x^{(i)}, z^{(i)}) = \log \sum_z Q^{(i)}(z) \underbrace{\cancel{\frac{P(x^{(i)}, z^{(i)})}{Q^{(i)}(z)}}}_{C}$

$$= \log C$$

Restate EM

(E-STEP) for $i = 1..n$, SET $Q_i(z) = P(z \mid x^{(i)}; \theta^{(t)})$

(M-STEP) $\theta^{(t+1)} = \arg\max_\theta L_t(\theta)$

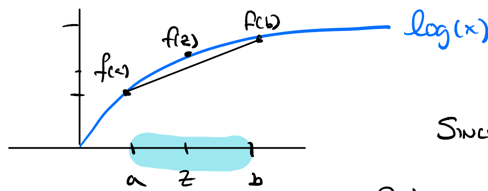
$$= \arg\max_\theta \sum_{i=1}^{n} \mathrm{ELBO}(x^{(i)}, Q^{(i)}, \theta)$$

Jensen Reminder (Sidebar)

$$\log(\mathbb{E}[x]) \geq \mathbb{E}[\log(x)]$$

log is concave function

To help you remember inequality, thought experiment $x \to a$ w/ prob λ, $x \to b$ w/ prob $1-\lambda$



Since z is like $\mathbb{E}[x] = \lambda a + (1-\lambda) b$

Picture says $f(\mathbb{E}[z]) \geq \mathbb{E}[f(z)]$ for $f = \log$.

for any $z = \lambda a + (1-\lambda) b$