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

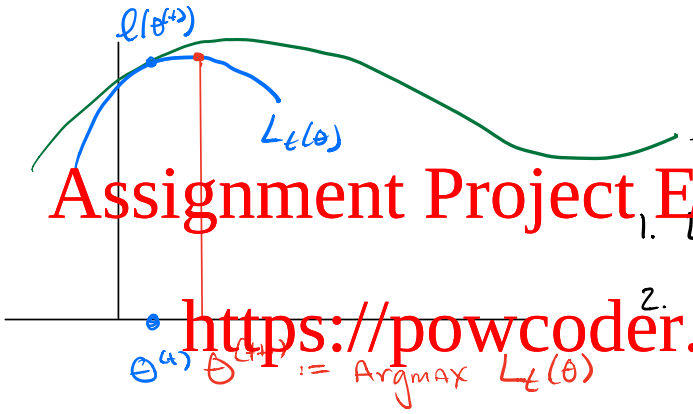
GUESS $z^{(i)}$ ↩ ↪ FIT PARAMETERS OF MODEL

Latent variable model

$$\ell(\theta) = \sum_{i=1}^{n} \log P(x^{(i)}; \theta) = \sum_{i=1}^{n} \sum_{z} \log P(x^{(i)} | z^{(i)} = z; \theta) P(z^{(i)} = z; \theta)$$

(data; parameters; latent variable)

Picture of our Algorithm (cf w/ Gmm)



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

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 pushy! 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)

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

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

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

So PICK ANY $Q^{(i)}$ → GIVES AN $L_i(\theta)$!

CALL THIS $ELBO(x, Q;\theta) = \sum_z Q(z) \log \frac{P(x,z;\theta)}{Q(z)}$

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

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

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

(cancelled term $= 1$; $P(x^{(i)}; \theta^{(t)})$ does not depend on z $= C$)

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

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

$$= \log C$$

Restate EM

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

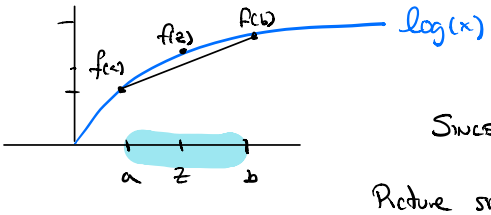
(M-STEP) $\theta^{(t+1)} = \underset{\theta}{\text{Argmax}}\ L_t(\theta)$

$$= \underset{\theta}{\text{Argmax}} \sum_{i=1}^{n} \text{ELBO}(x^{(i)}, Q^{(i)}, \theta)$$

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