→ Presentation slot : please check
→ Dec 7^{th} (Final project report - 5 pages)
see Piazza note

→ Dec 1^{st} HW4.

Generative models / Sampling

→ LDA, GMM
→ ML estimates
• EM algorithm
] Last time

→ EM algorithm ; application to GMMs
→ Variational Inference (VAEs implement approximate inference)

→ Indirect sampling : MCMC
Markov Chain Monte Carlo

(Not cover : adaptive MCMC)

Goal:
→ Score Generative model / Diffusion model

Distinction b/w inference methods, sampling algorithms & Generative modeling

Inference via sampling:
Target $\underset{\uparrow}{\pi(z)}$
Latent variable

Input: $\pi(z) = q(z|x)$

Generative assumption $P_\theta(x, z) = P_\theta(z) \underbrace{P_\theta(x|z)}_{\text{likelihood}}$

VI : $\underset{q \in Q}{\operatorname{argmin}} D_{KL}(q \,\|\, \underset{\pi(z)}{P_\theta(z|x)})$

→ Mean-field assumption

→ VAE (Lecture 18)

VI : approximate Bayesian inference

Sampling : (No Latent variable models / beyond inference)
Want: samples from π

Input: Unnormalized π is available

$$\pi(x) = \frac{e^{-V(x)}}{\underset{\uparrow}{Z}}$$
Normalization constant

$$\left(P_\theta(z|x) = \frac{P_\theta(x, z)}{\boxed{P_\theta(x)}} \right)$$

$e^{-V(x)}$
$-V(x)$

Input: $V(x)$
$\text{Score}_\pi(x) =$
$$\nabla \underset{\uparrow}{\log \pi}(x) = -\nabla V(x)$$
probability density

$x \in \mathbb{R}^d$
$\text{Score}_\pi(x) \in \mathbb{R}^d$
$V: \mathcal{X} \to \mathbb{R}$
$\text{Score}_\pi: \mathcal{X} \to \mathbb{R}^d$

$e^{-V(x)}$

- Transport-based sampling algorithms:
  - Triangular transport
  - Optimal transport
  - Normalizing flows

- Particle-based sampling
  - Optimal transport
  - Langevin dynamics & variants
  - MCMC variants

Generative model
Want: samples from π

Input: some samples from π

Score-generative models: learn score + sampling

Expectation maximization convergence

Iterative algorithm:

E-step: $$q_{t+1,i}(z) = P_{\theta_t}(z \mid x_i)$$

M-step: $$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{m} ELBO(q_{t+1,i}, \theta, x_i)$$

$$ELBO(q, \theta, x) = -\underset{q}{E} \log q + \underset{q}{E} \log P_\theta(x, z)$$

EM algorithm increases the likelihood at every step

- $$\log P_{\theta_{t+1}}(x) = l(\theta_{t+1}, x) \geq l(\theta_t, x)$$

- $$\begin{aligned} l(\theta_{t+1}, x) &= ELBO(q_{t+2}, \theta_{t+1}, x) \\ &\geq ELBO(q_{t+1}, \theta_{t+1}, x) \\ &\overset{(M\text{-step})}{\geq} ELBO(q_{t+1}, \theta_t, x) \\ &= l(\theta_t, x) \end{aligned}$$

$$P_\theta(\cdot \mid x) = \underset{q}{\operatorname{argmax}}\ ELBO(q, \theta, x) \quad (\text{E-step})$$

Gaussian Mixture model

Generative assumption / Probalistic model

$$P_\theta(X, Z) = \sum_{j=1}^{k} \frac{\pi_j \, e^{-(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)}}{\mathcal{Z}}$$

θ: set of parameters

X: observed

Z: Latent variable

$\mathcal{Z}$: normalization constant

$X \in \mathbb{R}^d$

$Z \in [k]$

$P(Z=j) = \pi_j$

k : no of components

$$\theta = (\underset{\mathbb{R}^d}{\mu_1}, \ldots, \mu_k, \underset{\mathbb{R}^{d \times d}}{\Sigma_1}, \ldots, \Sigma_k, \pi_1 \ldots \pi_k)$$

Z: Latent variables

$$P_\theta(X \mid Z=j) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-(x-\mu_j)^T \Sigma^{-1} (x-\mu_j)}$$

Z: clustering assignment

ML estimation using EM for GMM

$\theta = (\pi_1, \dots, \pi_k, \mu_1, \dots, \mu_k)$

E-step $\quad \theta_t$

$\rightarrow \quad q_{t+1}(z=j/x) = p_{\theta_t}(z/x)$

$= p_{\theta_t}(z=j) \cdot p_{\theta_t}(x/z=j)$

$$= \pi_j \times \frac{e^{-\frac{1}{2}(x-\mu_j)^T \Sigma^{-1}(x-\mu_j)}}{(2\pi)^{d/2} \, |\Sigma|^{1/2}}$$

$\mu_j \equiv \mu_{j,t} \quad \pi_j \equiv \pi_{j,t}$

$\rightarrow$ M step

$$\underset{\substack{\theta \\ (\pi_1, \dots, \pi_k, \mu_1, \dots, \mu_k)}}{\arg\max} \; ELBO(q_{t+1}, \theta)$$

$(- \mathbb{E}_q \log q + \mathbb{E}_q \log p_\theta(x,z))$

$$ELBO(q_{t+1}, \theta) = \sum_{i=1}^{m} \sum_{j=1}^{k} q_{t+1}(j/x_i) \log p_\theta(x_i, z=j)$$

$$\log \pi_j - \frac{1}{2}(x_i - \mu_j)^T \Sigma^{-1} (x_i - \mu_j)$$

$$\left(\sum_{j=1}^{k} \pi_j = 1 \right)$$

$$ELBO(q_{t+1}, \{\pi_j, \mu_j\}_{j=1}^{k}) + \lambda \left(\sum_{j=1}^{k} \pi_j - 1 \right)$$

GMM

$$\rightarrow \quad \mu_{j,t+1} = \sum_{i=1}^{m} \overbrace{P_{\theta_t}(z=j \mid X=x_i)}^{q_{t+1}} \, x_i$$

$$\pi_{j,t+1} = \frac{\sum_{i=1}^{m} \overbrace{P_{\theta_t}(z=j \mid X=x_i)}^{q_{t+1}}}{Z}$$

Space-shuttle O-ring model

Dalal et al 1989, A. Gandy lecture notes

→ Observations of failure from tests at various temperatures

$$P_\theta(x/z=1) = \frac{e^{\alpha} e^{x\beta}}{1+e^{\alpha}e^{x\beta}}$$

x : temperature

θ : (α, β)

→ Prior on θ

$$P(\alpha,\beta) = \frac{1}{b} e^{\alpha} e^{-e^{\alpha}/b}$$

Uniform on β.

→ $$P_\theta(z|x) = \frac{\boxed{P_\theta(z,x)}}{\boxed{P_\theta(x)}}$$

$$P_\theta(x) = \int P_\theta(x|z)\, P_\theta(z)\, dz$$

VI

$$q(\mu, \pi) = \prod_{j=1}^{k} \mathcal{N}(\mu_k ; m_k, s_k^2) \prod_{j=1}^{k} \pi_j$$

$$\mathrm{ELBO}(q, \theta, x)$$

$$= \sum_{j=1}^{k} \mathbb{E} \log \mathcal{N}(\mu_k ; m_k, s_k^2)$$

$$+ \sum_{j=1}^{k} \underline{\mathbb{E}} \log \pi_j + \mathbb{E} \log p_\theta(x|z)$$

$$- \mathbb{E}_q \log q$$

Markov chain Monte Carlo:

Construct a Markov Chain whose
→ invariant / stationary measure is the target π.

→ $\pi = e^{-V}/Z$ Have: $e^{-V(x)}$ $x_i \in \mathbb{R}^d$ A

→ Markov Chain: $X_1, \ldots, X_n \cdots$

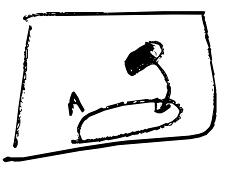
$P(X_{t+1} \mid X_t, \ldots, X_1) = P(X_{t+1} \mid X_t)$

→ $K(x, A) = P(X_{t+1} \in A \mid X_t = x)$

→ (X_n) is Harris recurrent if $\exists$ a measure π s.t. $\forall A$ with $\pi(A) > 0 \Rightarrow X_n$ visits A infinitely often

$\pi: \mathbb{R}^d \to \mathbb{R}^+$ $\pi(A) = P(X \in A)$

→ $$\pi(A) = \int_{\mathcal{X}} \underbrace{K(x, A)}\ \underbrace{\pi(dx)}$$



$$\sum_{X_t} \pi(X_t) \cdot P(X_{t+1} = y \mid X_t) = \pi(y)$$

→ Ergodic theorem:

$f \in L^1(\pi)$ then

$$\frac{1}{n} \sum_{i=1}^{n} f(X_i) \to \int f(x) \pi(dx) = \mathbb{E}_{X \sim \pi} f(X)$$

→ CLT for time-inhomogeneous Markov chain with invariant measure π

- Let $g \in L^1(\pi)$ with $Var(g(X)) < \infty$
- $$\hat{g}_N = \frac{1}{N} \sum_{n=0}^{N-1} g(X_n)$$
- $$\sigma^2 := \lim_{N \to \infty} Var(\sqrt{N}\, \hat{g}_N) = Var(g(X)) + 2 \sum_{n=1}^{\infty} cov(g(X_1), g(X_{1+n}))$$
- $$(\hat{g}_N - \mathbb{E}_{X \sim \pi} g(X)) \sqrt{N} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Convergence to stationary distribution

For Harris recurrent, aperiodic Markov chain with invariant measure π,

$$\lim_{n \to \infty} \left\| \int K^n(x, \cdot)\, \mu(dx) - \pi \right\|_{TV} = 0$$

$$\| \mu_1 - \mu_2 \|_{TV} = \sup_A | \mu_1(A) - \mu_2(A) |$$