Session 7:

1403, 12, 19

لکچر نوت های D.M. Blei در Princeton خوبه.

$$P(\theta) \quad \text{which} \quad \theta = (\mu, \sigma, \dots)$$

$$\hat{x} = \hat{D}$$

$$P(\theta \mid x) = \frac{P(x \mid \theta)\, P(\theta)}{P(x)}$$

likelihood با Maximum likelihood متفاوت است.

در اینجا $P(x \mid \theta)$، likelihood است.

$$P(x) = \int_\theta P(x, \theta)\, d\theta = \int_\theta P(x \mid \theta)\, P(\theta)\, d\theta$$

به دست می‌آید.

پارامتر های پنهان

$$Z^m = \hat{\theta}$$

$$P(Z^m \mid x^n) = \frac{P(Z^m, x^n)}{P(x^n)}$$

$$q(Z^n)$$

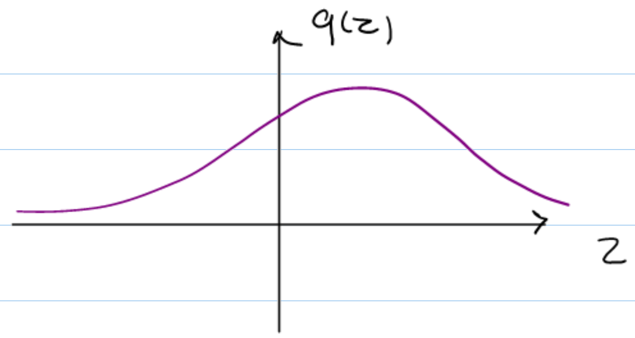
$$q^*(Z^m) = \arg\max \left[ELBO \right]$$

Which

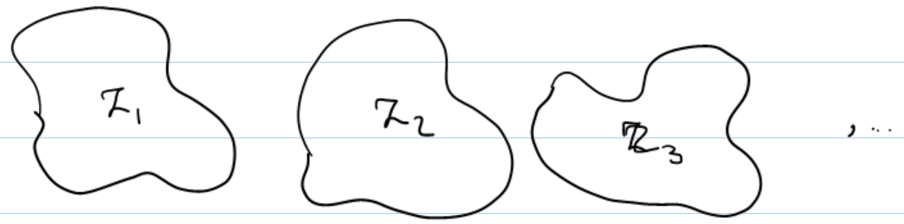
$$ELBO = E_{q(Z^m)} \left[\log \left\{ \frac{P(Z^m, x^n)}{q(Z^m)} \right\} \right]$$

در اینجا به زبان مطالب تحلیلی Z^m زمان ما است و q پارامتر مطالب است.

$$Z^m = t \quad , \quad q = Q \quad \Rightarrow Q(t)$$



مثال) یک hidden variable داریم که در ذهن ما متوسط یک توزیع است

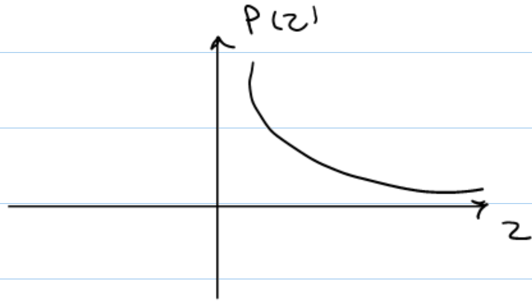


$$Z \propto Exp(z; \lambda = 1)$$

$$\Rightarrow P(Z) = Exp(Z; \lambda = 1) = \begin{cases} \lambda e^{-\lambda Z} & Z \geq 0 \\ 0 & \text{else} \end{cases}$$

$$x \sim \mathcal{N}(x, \mu = z, \sigma = 1)$$

هر zای که برمی‌داریم از sample ها می‌سازیم.



$$= e^{-z} \underbrace{I(\underbrace{z \geq 0}_{1})}_{0}$$

در رابطه Bayesian داشتیم:

$$P(Z^m | x^n) , P(Z^m, x^n) \longrightarrow P(x^n | Z^m) P(Z^m)$$

μ را condition کرده‌ایم.

$$P(x|z) = \mathcal{N}(x; \mu = z, \sigma = 1)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-z)^2\right)$$

$$P(x,z) = P(z)\, P(x|z)$$

$$= e^{-z} I(z \geq 0) \frac{1}{2\pi} e^{-\frac{1}{2}(x-z)^2}$$

$P(x)$

$$P(z|x) = \frac{P(x,z)}{P(x)}$$ → Normalization ضریب

$$P(x) = \int_0^{\infty} e^{-z} I(z \geq 0) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-z)^2} dz$$

این ها انتگرال هستند ⇐ intractable

برای آنکه حوال زده نشویم فقط 1 بار آزمایش می‌کنیم ⇐ 1 رادیو می‌خرم.

$$P(z|x) = e^{-z} I(z \geq 0) e^{-\frac{1}{2}(x-z)^2}$$

fix

$$= \exp\left[-z - \cancel{\frac{1}{2}x^2} + xz - \frac{1}{2}z^2\right] I(z \geq 0)$$

fix

$$\sim \exp\left\{-\frac{1}{2}z^2 + (x-1)z\right\} I(z \geq 0)$$

باید اندازه بری تبدیل شد به Gaussian ...!

اگر $\frac{1}{2}$ رادیو همون Variance را داره که هم دارند تا می‌توان Variance را هم

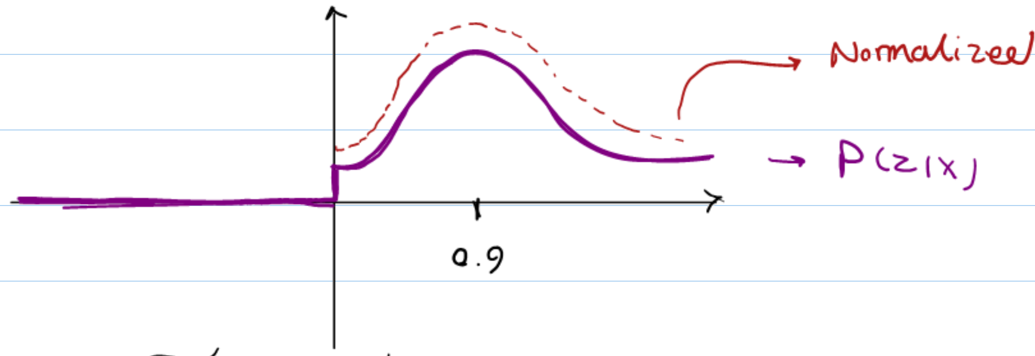
پیدا کرد.

$$\sim \exp\left[-\frac{1}{2}(z-(x-1))^2 + \frac{1}{2}(x-1)^2\right] I(z \geq 0)$$

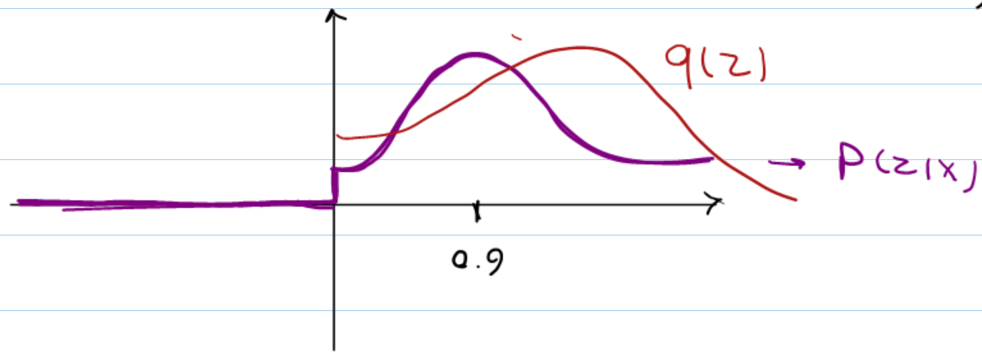
$$\sim \exp\left[-\frac{1}{2}\big(z - \underbrace{(x-1)}\big)^2\right] \; I(z \geq 0)$$

که به سمت μ میرود خودش نه به -1 است.

plot این تابع توزیع $\qquad x_0 = 1.9$



حال $P(z|x)$ را $q(z)$ تخمین بزنیم که فاصله آنها min باشد یا ELBO max باشد



$$q^*(z) = \underset{q(z)}{\arg\max}\,[\text{ELBO}]$$

$$\text{ELBO} = E_{\substack{q(z^m) \\ m=1}}\left[\log \frac{P(z^m, x=D)}{q(z^m)}\right]$$

تا اس

حال { پارامتریک . از توزیع شروع می‌کنیم و می‌خواهیم تقریب‌هایمان کنیم ← مسئله بیزی

غیر پارامتریک : هیچ ایده‌ای نداریم $q(z)$ چیست ← مانند معادله اویلر-لاگرانژ

$$q_\theta(z) = \text{Exp}\{ z , \lambda = \theta \}$$

شکل همان می‌ماند، پارامتر تعیین می‌شد.

$$= \begin{cases} \theta e^{-\theta z} & z \geq 0 \\ 0 & \text{else} \end{cases}$$

$$= \theta e^{-\theta z} \, I(z \geq 0)$$

$$\theta^* = ? = \arg\max \; E_{q(z)} \left[\log \frac{P(z, x = D)}{q_\theta(x)} \right]$$

$$\text{ELBO} = E_{q_\theta(z)} \left[\log P(z, x = D) - \log(q_\theta(z)) \right]$$

$$= E_{q_\theta(z)} \Big[-z + \cancel{\log(I(z \geq 0))}^{\,0} - \frac{1}{2}\log(2\pi) - \frac{1}{2}(x_0 - z)^2 - (\log\theta - \theta z + 0) \Big]$$

$$= \mathbb{E}_{q_\theta(z)}\left[-z - \frac{1}{2}x_0^2 + x_0 z - \frac{1}{2}z^2 - \log\theta + \theta z \right]$$

$$\mathbb{E}_{q_\theta(z)}[\cdots] = \cdots \quad , \quad \mathbb{E}_{q_\theta(z)}[\cdots z] = \frac{1}{\theta} \cdots$$

$$\mathbb{E}_{q_\theta(z)}[\cdots , z^2] = \frac{2}{\theta^2} \cdots$$

$$\text{ELBO} = -\log\theta + \frac{x_0 - 1}{\theta} - \frac{1}{\theta^2}$$

$$\frac{\partial}{\partial} \Rightarrow \theta^* \quad \rightarrow \quad -\frac{1}{\theta} + \frac{1}{\theta^2} - \frac{x_0}{\theta^2} + \frac{2}{\theta^3} = 0$$

$$\theta^2 + (x_0 - 1)\theta - 2 = 0$$

$$\theta_{1,2} = \frac{-(x_0 - 1)}{2} \pm \sqrt{\frac{(\quad)^2}{4} + 2}$$

مثبت را می‌خواهیم.

$$\Rightarrow \theta^* = -\frac{x_0 - 1}{2} + \sqrt{\frac{(x_0 - 1)^2}{4} + 2}$$

کل این جلسه نشان می‌دهد که پاسخ کامپیوتر بسیار غلط است

تنها 2 ابزار برای این داریم

ELBO و Cavi را در این خیلی زیاد می‌شنویم

Cavi برای توابع Convex استفاده می‌شود ← min راحت دارند

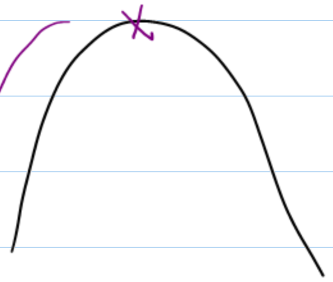
Coordinate Ascent Maximization

مسئله m بعدی که opt انجام دهیم ← z_1 را به عنوان متغیر بگیریم و بقیه را fix

کنیم و نسبت به آن ELBO را Max می‌کنیم ← مقدار می‌دهد.

بعد می‌رویم سراغ z_2 و تکرار ← یک جایی z^* از جواب‌ها به دست

می‌آید.



این را حتماً پیدا می‌کند.

این تضمین نیست که در حقیقت توابع $\cap$ به این شکل باشند.

روش دیگر که برای 2بعدی است:

Coordinate Ascent Meanfield variational Inference

Cavi

در این opt و θ داشتیم ولی می‌توانستیم چندتا داشته باشیم اما این ها هم ربطی ندارند ← var به mean ارتباطی ندارد.

Mean-Field :

$$q(z^m) = \prod_{i=1}^{m} q(z_i)$$

مثلاً برای Gaussian تک var و μ فرضیات اگر یکی امکان دارد Moment چندم داشته باشیم که جالب نیست چون از Moment دوم بدست میاید.

$$ELBO = E_{q(z^m)}\left[\log \frac{p(z^m, x^n)}{q(z^m)}\right]$$

$$= \int q(z^m) \log\left[\frac{p(z^m, x^n)}{q(z^m)}\right] dz^m$$

$$q^*(z^m) = \underset{q(z^m)}{\arg\max}\left[ELBO\right]$$

$$ELBO = \log(p(x^n)) + E_{q(z^m)}\left[\log(p(z^m|x^n))\right] - E_{q(z^m)}\left[\log q(z^m)\right]$$

$$= \text{Const} + E_{q(z^m)}\left[\log p(z^m \mid x^n,\right]$$

$$- \sum_{i=1}^{m} E_{q(z^m)}\left[\log q(z_i)\right]$$

$$P(z_i, z^{-i} \mid x^n) = P(z_i \mid z^{-i}, x^n)\, P(z^{-i} \mid x^n)$$

$$E_{q(z^{-i})} \log p(z_i, z^{-i} \mid x^n)$$

$$= E_{q(z^{-i})}\left[\log p(z_i \mid z^{-i}, x^n)\right]$$

$$+ E_{q(z^{-i})}\left[\log p(z^{-i} \mid x^n)\right] \simeq 0$$

امکان داره در ادبیات

این دو را مساوی هم

بیتم خیلی غیر منطقی نیست!

$$q^*(z_i) = \underset{q(z^m)}{\arg\max}\ \text{ELBO}$$

$$= \arg\max\ E_{q(z_i)} E_{q(z^{-i})}\left[\log p(z_i \mid z^{-i}, x^n)\right.$$

$$\left. - E_{q(z_i)}\left(\log q(z_i)\right)\right.$$

$$\frac{\delta \text{ELBO}}{\delta q(z_i)} = E_{q(z^{-i})}\left[\log p(z_i \mid z^{-i}, x^n)\right]$$

$$- \ln q^*(z_i) + 1 = 0$$

$$q^*(z_i) = C\, e^{E_{q(z^{-i})}\left[\log p(z_i \mid z^{-i}, x^n)\right]}$$