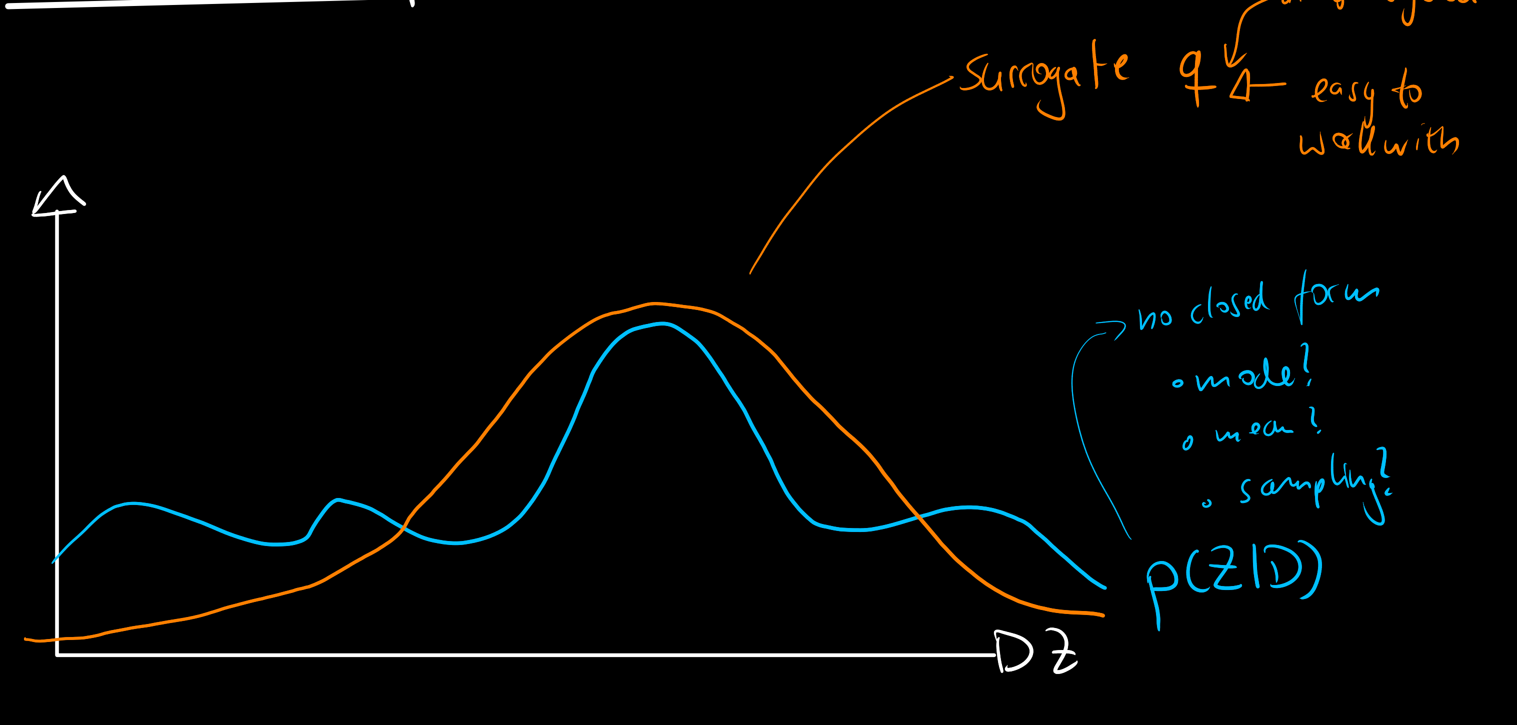
Variational Inference & ELBO



DGM with observed $\underline{X}_{\mathbb{R}^N}$ and latent $\underline{Z}_{\mathbb{R}^D}$

different sizes

i.e. we have the joint $p(\underline{X}, \underline{Z})$

"inference": we observe $\underline{X}$ as Dataset

↳ we want the posterior

Bayes' Rule: $p(\underline{Z} \mid \underline{X}=D) = \dfrac{p(\underline{X}=D \mid \underline{Z}) \cdot p(\underline{Z})}{p(\underline{X}=D)}$

problem: marginal

$$p(\underline{X}) = \int_{z_0} \cdots \int_{z_{D-1}} p(\underline{X}, \underline{Z}) \, dz_0 \cdots dz_{D-1}$$

intractable ($\hat{=}$ incomputable) in general

Remedy

instead find $q(\underline{Z}) \approx p(\underline{Z} \mid \underline{X}=D)$ as good as possible

"variational": optimize for a function

Goodness of the fit?

→ KL-Divergence $\hat{=}$ "distance between two dist"

! minimize KL

$$q^*(\underline{Z}) = \underset{q(\underline{Z}) \in Q}{\arg\min} \left(KL\left(q(\underline{Z}) \,\|\, \underbrace{p(\underline{Z} \mid \underline{X}=D)}_{\hat{=}\, p(\underline{Z}|D)} \right) \right)$$

a family of "simple" distributions

$$KL(q(\underline{Z}) \,\|\, p(\underline{Z}|D)) = \mathbb{E}_{\underline{Z} \sim q(\underline{Z})} \left[\log \frac{q(\underline{Z})}{p(\underline{Z}|D)} \right]$$

$$= \int_{z_0} \cdots \int_{z_{D-1}} q(\underline{Z}) \cdot \log \frac{q(\underline{Z})}{p(\underline{Z}|D)} \, dz_0 \cdots dz_{D-1}$$

but we don't have the posterior :(

we only have the joint $p(\underline{Z}, D)$

Remedy

Rearrange

$$p(\underline{Z}|D) = \frac{p(\underline{Z}, D)}{p(D)}$$

$$KL(q(\underline{Z}) \,\|\, p(\underline{Z}|D)) =$$

$$= \int_{\underline{Z}} q(\underline{Z}) \log \left(\frac{q(\underline{Z})}{p(\underline{Z}|D)} \right) d\underline{Z}$$

$$= \int_{\underline{Z}} q(\underline{Z}) \cdot \log \left(\frac{q(\underline{Z}) \cdot p(D)}{p(\underline{Z}, D)} \right) d\underline{Z}$$

$$= \int_{\underline{Z}} q(\underline{Z}) \cdot \log \left(\frac{q(\underline{Z})}{p(\underline{Z}, D)} \right) d\underline{Z} + \int_{\underline{Z}} q(\underline{Z}) \log(p(D)) \, d\underline{Z}$$

$$= \mathbb{E}_{\underline{Z} \sim q(\underline{Z})} \left[\log \left(\frac{q(\underline{Z})}{p(\underline{Z}, D)} \right) \right] + \mathbb{E}_{\underline{Z} \sim q(\underline{Z})} \left[\log p(D) \right]$$

$$= \underbrace{- \mathbb{E}_{\underline{Z} \sim q(\underline{Z})} \left[\log \left(\frac{p(\underline{Z}, D)}{q(\underline{Z})} \right) \right]}_{\mathcal{L}(q)} + \log p(D)$$

$$\underbrace{KL}_{\text{something positive}} = - \underbrace{\mathcal{L}(q)}_{\substack{\text{something negative} \\ \leq 0}} + \underbrace{\log \underbrace{p(D)}_{\text{marginal}}}_{\text{evidence}}$$

→ distance between posterior & surrogate

→ smaller than $\log p(D)$

→ Evidence Lower Bound (ELBO)

→ something negative ≤ 0

! fixed

$$\boxed{\text{ELBO} \quad \mathcal{L}(q) = \mathbb{E}_{\underline{Z} \sim q(\underline{Z})} \left[\log \frac{p(\underline{Z}, D)}{q(\underline{Z})} \right]}$$

$\mathcal{L}(q) = \log p(D)$ i.f.f. $KL(q(\underline{Z}) \,\|\, p(\underline{Z}, D)) = 0$

→ usually not achieved in VI

Variational inference

$$q^*(\underline{Z}) = \underset{q(\underline{Z}) \in Q}{\arg\min} \left(KL\left(q(\underline{Z}) \,\|\, p(\underline{Z}|D) \right) \right)$$

equivalent

$$q^*(\underline{Z}) = \underset{q(\underline{Z}) \in Q}{\arg\max} \left(\mathcal{L}(q) \right)$$

$$KL = -\mathcal{L} + \log p$$

$\Leftarrow$ $\boxed{\mathcal{L}(q) = - KL + \log p(D)}$