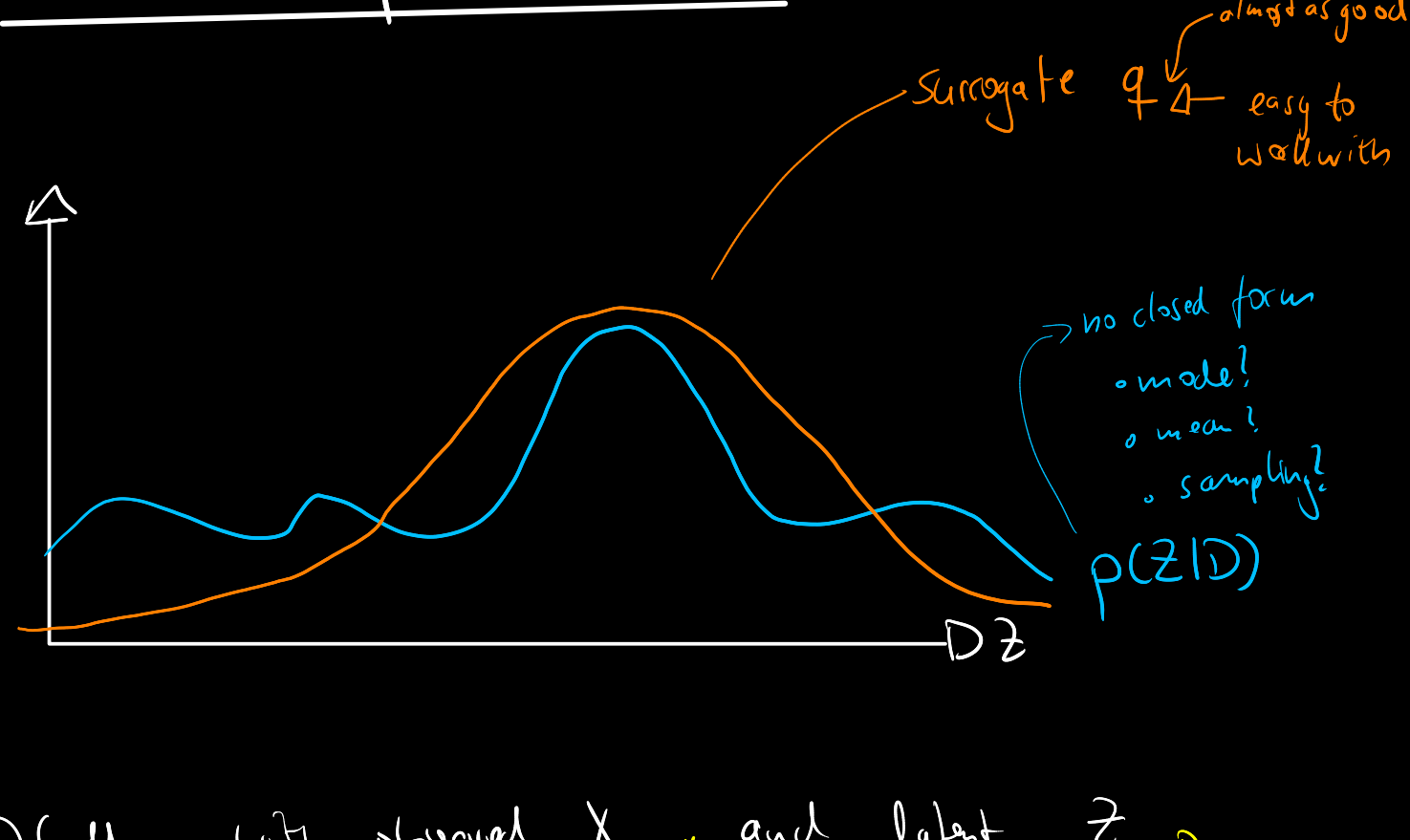


Variational Inference & ELBO



DL with observed $\underline{x} \in \mathbb{R}^n$ and latent $\underline{z} \in \mathbb{R}^D$

└──────────────────┘
different sites

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

"inference": we observe \underline{x} as Dataset

↳ we want the posterior

Bayes' Rule: $p(\underline{z} | \underline{x} = D) = \frac{p(\underline{x} = D | \underline{z}) \cdot p(\underline{z})}{p(\underline{x} = D)}$

problem: marginal

$$p(\underline{x}) = \int \dots \int p(\underline{x}, \underline{z}) d z_0 \dots d z_{D-1}$$

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

Remedy

instead find $q(\underline{z}) \approx p(\underline{z} | \underline{x} = D)$
as good as possible

"Variational": optimize for a function

Goodness of the fit?

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

∇ minimize KL

$$q^*(\underline{z}) = \arg \min_{q(\underline{z}) \in \mathcal{Q}} (KL(q(\underline{z}) || \underbrace{p(\underline{z} | \underline{x} = D))}_{\hat{=} p(\underline{z} | D)})$$

↑
a family of
"simple" distributions

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

$$= \int \dots \int q(\underline{z}) \cdot \log \frac{q(\underline{z})}{p(\underline{z} | D)} d z_0 \dots d z_{D-1}$$

but we don't have the posterior (∴)

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

Remedy

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

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

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

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

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

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

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

something positive something negative ≤ 0 evidence

→ distance between posterior & surrogate → smaller than $\log p(D)$ → something negative ≤ 0

→ Evidence Lower Bound (ELBO) ∇ find

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

$\mathcal{L}(q) = \log p(D)$ i.f.f. $KL(q(\underline{z}) || p(\underline{z}, D)) = 0$
→ usually not achieved in VI

Variational Inference

$$q^*(\underline{z}) = \arg \min_{q(\underline{z}) \in \mathcal{Q}} (KL(q(\underline{z}) || p(\underline{z}, D)))$$

equivalent

$$q^*(\underline{z}) = \arg \max_{q(\underline{z}) \in \mathcal{Q}} (\mathcal{L}(q))$$

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

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