1 Generative modelling

Learn $p_{\text{model}} \approx p_{\text{data}}$, sample from p_{model}.

- Explicit density:
  - Approximate:
    - Variational: VAE, Diffusion
    - Markov Chain: Boltzmann machine
  - Tractable:
    - Autoregressive: FVSBN/NADE/MADE, Pixel(c/r)nn, WaveNet/tcn, Autor. Transf.,
    - Normalizing Flows
- Implicit density:
  - Direct: Generative Adversarial Networks
  - MC: Generative Stochastic Networks

Autoencoder: $X \to Z \to X$, $g \circ f \approx \text{id}$, f and g are NNs. Optimal linear autoencoder is PCA. Undercomplete: $|Z| < |X|$, else overcomplete. Overcomp. is for denoising, inpainting. Latent space should be continuious and interpolable. Autoencoder spaces are neither, so they are only good for reconstruction.

2 Variational AutoEncoder (VAE)

Sample z from prior $p_\theta(z)$, to decode use conditional $p_\theta(x \mid z)$ defined by a NN.

$D_{\text{KL}}(P\|Q) \coloneqq \int_x p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$: KL divergence, measure similarity of prob. distr. $D_{\text{KL}}(P\|Q) \neq D_{\text{KL}}(Q\|P)$, $D_{\text{KL}}(P\|Q) \geq 0$

Likelihood $p_\theta(x) = \int_z p_\theta(x \mid z) p_\theta(z) \mathrm{d}z$ is hard to maximize, let encoder NN define $q_\phi(z \mid x)$, $\log p_\theta(x^i) = \mathbb{E}_z \left[\log p_\theta(x^i \mid z)\right] - D_{\text{KL}}(q_\phi(z \mid x^i) \| p_\theta(z)) + D_{\text{KL}}(q_\phi(z \mid x^i) \| p_\theta(z \mid x^i))$. Red is intractable, use ≥ 0 to ignore it; Orange is reconstruction loss, clusters similar samples; Purple makes posterior close to prior, adds cont. and interp. Orange – Purple is **ELBO**, maximize it.

$x \xrightarrow{\text{enc}} \mu_{z|x}, \Sigma_{z|x} \xrightarrow{\text{sample}} z \xrightarrow{\text{dec}} \mu_{x|z}, \Sigma_{x|z} \xrightarrow{\text{sample}} \hat{x}$

Backprop through sample by reparametr.: $z = \mu + \sigma\epsilon$. For inference, use μ directly.

Disentanglement: features should correspond to distinct factors of variation. Can be done with semi-supervised learning by making z conditionally independent of given features y.

2.1 β-VAE

Disentangle by $\max_{\theta,\phi} \mathbb{E}_x \left[\mathbb{E}_{z \sim q_\phi} \log p_\theta(x \mid z)\right]$ s.t. $D_{\text{KL}}(q_\phi(z \mid x) \| p_\theta(z)) < \delta$, with KKT: max Orange $- \beta$Purple.