Information-Theoretic Diffusion [Kong et al]

Let $p(z_\gamma|x)$ be a Gaussian noise channel with $z_\gamma = \sqrt{\gamma}\, x + \varepsilon$

where $\varepsilon \sim \mathcal{N}(0, I)$, γ represents the signal-to-noise ratio, and $p(x)$ is unknown data distribution. Marginal is $p(z_\gamma) = \int p(z_\gamma|x)p(x)dx$

and pointwise MMSE is defined as

$$\mathrm{mmse}(x,\gamma) = \mathbb{E}_{p(z_\gamma|x)}\left[\| x - \hat{x}^*(z_\gamma, \gamma) \|_2^2 \right]$$

where $\hat{x}^*(z_\gamma, \gamma) = \underset{\hat{x}(z_\gamma,\gamma)}{\arg\min}\ \mathrm{mse}(\gamma) = \mathbb{E}_{p(x|z_\gamma)}[x]$, the conditional mean.

Then $$\frac{d}{d\gamma} D_{KL}\left[p(z_\gamma|x) \,\|\, p(z_\gamma) \right] = \frac{1}{2}\,\mathrm{mmse}(x,\gamma).$$

Note that taking expectation with respect to x on both sides recovers Guo's original I-MMSE relation.

Diffusion as thermodynamic integration

Derive expression for log-likelihood that resembles variational bound.

→ thermodynamic variational inference constructs a path connecting a tractable distribution to some target such that integrating over path recovers log likelihood for target model.

→ generally computationally tough, but in diffusion models, have Gaussian noise channel that transforms target distribution into standard normal, so can be easily sampled at intermediate points

Want to use fundamental theorem of calculus ("thermodynamic integration")

$$\int_{\gamma_0}^{\gamma_1} d\gamma \frac{d}{d\gamma} f(\gamma) = f(\gamma_1) - f(\gamma_0).$$

Consider for $f(x,\gamma) = D_{KL}\left[p(z_\gamma|x) \,\|\, p(z_\gamma)\right]$

$$\int_{\gamma_0}^{\gamma_1} d\gamma \frac{d}{d\gamma} f(x,y) = D_{KL}\left[p(z_{\gamma_1}|x) \,\|\, p(z_{\gamma_1})\right] - D_{KL}\left[p(z_{\gamma_0}|x) \,\|\, p(z_{\gamma_0})\right]$$

$$= D_{KL}\left[p(z_{\gamma_1}|x) \,\|\, p(z_{\gamma_1})\right] - \mathbb{E}_{p(z_{\gamma_0}|x)}\left[\log p(x|z_{\gamma_0})\right] + \log p(x)$$

by expanding KL divergence and using Bayes rule.

Rearranging:

$$-\log p(x) = D_{KL}\left[p(z_{\gamma_1}|x) \,\|\, p(z_{\gamma_1})\right] + \mathbb{E}_{p(z_{\gamma_0}|x)}\left[\log p(x|z_{\gamma_0})\right] - \int_{\gamma_0}^{\gamma_1} d\gamma \frac{d}{d\gamma} f(x,y)$$

now use pointwise I-MMSE to get

$$-\log p(x) = \underbrace{D_{KL}\left[p(z_{\gamma_1}|x) \,\|\, p(z_{\gamma_1})\right]}_{\text{prior loss}} + \underbrace{\mathbb{E}_{p(z_{\gamma_0}|x)}\left[\log p(x|z_{\gamma_0})\right]}_{\text{reconstruction loss}} \underbrace{- \frac{1}{2}\int_{\gamma_0}^{\gamma_1} \mathrm{mmse}(x,\gamma)\, d\gamma}_{\text{diffusion loss}}$$

Can we get even simpler (exact) expression, essentially by looking at non-Gaussianity.

$$-\log p(x) = \frac{d}{2}\log(2\pi e) - \frac{1}{2}\int_0^{\infty} d\gamma \left(\frac{d}{1+\gamma} - \mathrm{mmse}(x,\gamma)\right)$$

density written solely in terms of global optimum of a particular regression problem: denoising MSE.

→ neural networks good at unconstrained optimization of MSE loss functions

$h(p)$ differential privacy is expected value:

$$h(p) = \mathbb{E}_{p(x)}\left[-\log p(x)\right] = \underbrace{\frac{d}{2}\log 2\pi e}_{\text{Gaussian entropy}} - \underbrace{\frac{1}{2}\int_0^{\infty} d\gamma \left(\frac{d}{1+\gamma} - \text{mmse}(\gamma)\right)}_{\text{deviation from Gaussianity } \geq 0.}$$

so focus on parts that deviate most from Gaussianity.

Also works as discrete probability estimator

$$-\log P(x) = \frac{1}{2}\int_0^{\infty} \text{mmse}(x,\gamma)\, d\gamma.$$