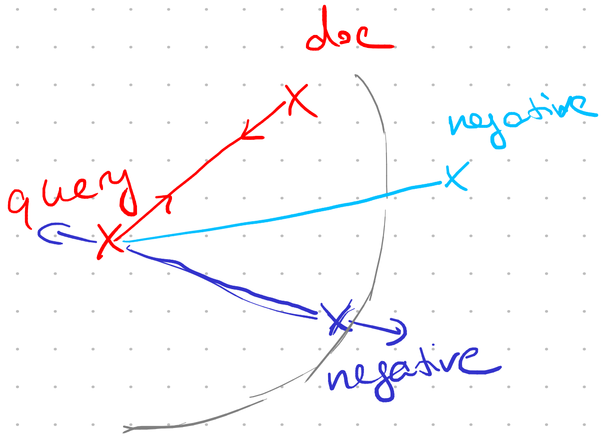
Triplet loss



hard negative mining

Θ

Explicit   Implicit

Tractable density

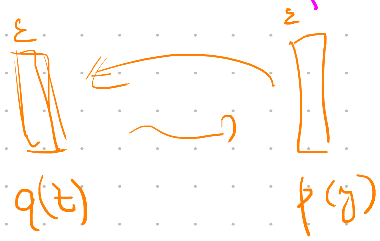
$p(x_i \mid x_1 \dots x_{i-1})$

$\cancel{\text{Gaussian}} \longrightarrow f_1 \longrightarrow f_2 \longrightarrow \dots \longrightarrow f_N \longrightarrow p(\bar{x})$

$p(\bar{z}) = \mathcal{N}(\bar{0}, I)$

$\bar{z} \sim q(\bar{z}) \xrightarrow{f} \bar{y} = \bar{f}(\bar{z})$, $\quad \xleftarrow{f^{-1}}$

$$p(\bar{y}) = p(\bar{f}(\bar{z})) = q(\bar{f}^{-1}(\bar{y})) \cdot \left|\det \frac{df^{-1}}{d\bar{y}}\right|$$

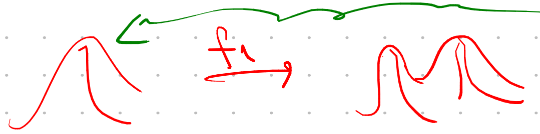


$$p(y) = q(z) \cdot \left|\frac{dz}{dy}\right| = q(f^{-1}(y)) \cdot \left|\frac{df^{-1}}{\partial y}\right| = q(f^{-1}(y)) \cdot f^{-1}{}'(y)$$

$\sim \mathcal{N}(z \mid \bar{0}, I)$

$$\underline{\bar{z} \sim q(\bar{z})} \rightsquigarrow f_\theta \rightsquigarrow \bar{y} \quad \underline{p(\bar{f}(\bar{z}))} = q(\bar{z}) \cdot \left|\det \frac{df}{dz}\right|^{-1} \quad \to \text{max}$$

$$\underline{\log p(\bar{f}(\bar{z}))} = \underline{\log q(\bar{z})} - \underbrace{\log\left|\det \frac{d\bar{f}}{d\bar{z}}\right|_\theta} \quad \xrightarrow{\text{max}} \bar{\theta}$$



$\xrightarrow{f_1} \quad \rightarrow \ldots \xrightarrow{f_K} \quad \bar{z}_K = f_K(\bar{z}_{K-1})$

$\bar{z}_0 \sim q(\bar{z}_0) \longrightarrow \bar{z}_1 \sim p_1(\bar{z}_1) \qquad \bar{z}_K \sim p_K(\bar{z}_K)$

$$p_k(\bar{z}_k) = p_k(f_k^{-1}(\bar{z}_{k-1})) \cdot \left|\det \frac{d f_k^{-1}}{d\bar{z}_k}\right| = \ldots$$

$$p_K(\bar{z}_K) = q(\bar{z}_0) \cdot \left|\det \frac{d f_1^{-1}}{d\bar{z}_1}\right| \left|\det \frac{d f_2^{-1}}{d\bar{z}_2}\right| \cdots \left|\det \frac{d f_K^{-1}}{d\bar{z}_K}\right|$$

$$\log p_K(\bar{z}_K) = \log q(\bar{z}_0) + \sum_{k=1}^{K} \log \left|\det \frac{d f_k^{-1}}{d\bar{z}_k}\right| \xrightarrow[\bar{\theta}]{} \max$$

$$= \log q(\bar{z}_0) - \sum_{k=1}^{K} \log \left|\det \frac{d \bar{f}_k}{d\bar{z}_{k-1}}\right|$$

$\bar{x}_0 \leftarrow \bar{z}_1 \leftarrow \ldots \leftarrow \bar{z}_{k-1} \leftarrow \bar{z}_k$

RealNVP

$$\begin{pmatrix} I & 0 \\ \ast & \ast \end{pmatrix}$$

$f: \bar{x} \longmapsto \bar{y} = (y_1 \ldots y_d \, y_{d+1} \ldots y_D)$

$(x_1 \ldots x_d \, x_{d+1} \ldots x_D)$

$$\begin{cases} \bar{y}_{1..d} = \bar{x}_{1..d} \\ \bar{y}_{d+1..D} = \bar{x}_{d+1..D} \odot e^{s(\bar{x}_{1..d})} + t(\bar{x}_{1..d}) \end{cases}$$

$\bar{x}_{1..d} = \bar{y}_{1..d}$

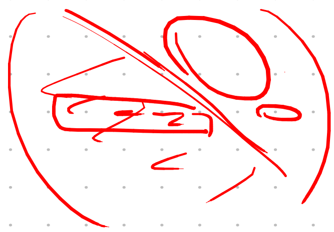
$\bar{x}_{d+1..D} = (\bar{y}_{d+1..D} - t(\bar{y}_{1..d})) \oslash e^{s(\bar{y}_{1..d})}$

$\bar{w}$

$$J = \begin{pmatrix} I & 0 \\ \frac{\partial \bar{y}_{d+1..D}}{\partial \bar{x}_{1..d}} & e^{s(\bar{x}_{1..d})} \end{pmatrix}$$

$$\det\left|\frac{df}{d\bar{x}}\right| = e^{\sum_{j=d+1}^{D} s_j(\bar{x}_{1..d})}$$

MAF – Masked Autoregressive Flows



$$p(\bar x) = \prod_{i=1}^{D} p(x_i \mid \bar x_{1..i-1}) \qquad \bar z \xmapsto{f} \bar x$$

$$x_i \sim p(x_i \mid \bar x_{1..i-1}) = z_i \cdot \sigma_i(\bar x_{1..i-1}) + \mu_i(\bar x_{1..i-1})$$

$$z_i = \frac{x_i - \mu_i(\bar x_{1..i-1})}{\sigma_i(\bar x_{1..i-1})} \qquad \boxed{\bar z = \frac{\bar x - \bar\mu(\bar x)}{\bar\sigma(\bar x)}}$$

$$\frac{df_i}{dz_j} = \frac{dx_i}{dz_j} \qquad \frac{dx_i}{dz_j} = 0, \; j>i$$

$$\frac{dx_i}{dz_j} = \ldots, \; j<i$$

$$\frac{dx_i}{dz_i} = \sigma_i(\bar x_{1..i-1}) \qquad \left|\det \frac{d\bar x}{d\bar z}\right| = \prod_i \sigma_i(\bar x_{1..i-1})$$

$f: \bar z \mapsto \bar x$ – autoregr. / SLOW

$f^{-1}: \bar x \mapsto \bar z$ – parallel / FAST

IAF – inverse autoregr. flows

$$x_i \sim p(x_i \mid \bar z_{1..i-1}) = z_i \cdot \sigma_i(\bar z_{1..i-1}) + \mu_i(\bar z_{1..i-1})$$

$$\bar x = \bar z \odot \bar\sigma(\bar z) + \bar\mu(\bar z)$$

$$z_i = \frac{x_i - \mu_i(\bar z_{1..i-1})}{\sigma_i(\bar z_{1..i-1})}$$

$f: \bar z \mapsto \bar x$ FAST

$f^{-1}: \bar x \mapsto \bar z$ SLOW

Diffusion-based models

$q(\bar{x}_0)$

$$\bar{x}_0 \to \bar{x}_1 \to \dots \to \bar{x}_t \to \dots \to \bar{x}_T$$

p_θ

$$q(\bar{x}_t|\bar{x}_{t-1}) = \mathcal{N}(\bar{x}_t|\sqrt{1-\beta_t}\cdot\bar{x}_{t-1}, \beta_t I)$$

$$q(\bar{x}_1 \dots \bar{x}_T|\bar{x}_0) = q(\bar{x}_1|\bar{x}_0) \dots q(\bar{x}_T|\bar{x}_{T-1})$$

$$q(\bar{x}_t|\bar{x}_0) = ?$$

$\sim \mathcal{N}(\bar{0}, I)$

$\alpha_t = 1-\beta_t$

$$\bar{x}_t = \sqrt{1-\beta_t}\cdot\bar{x}_{t-1} + \sqrt{\beta_t}\cdot\bar{\varepsilon}_{t-1} =$$

$$= \underbrace{\sqrt{1-\beta_t}}_{\alpha_t}\cdot\left(\underbrace{\sqrt{1-\beta_{t-1}}}_{\alpha_{t-1}}\cdot\bar{x}_{t-2} + \sqrt{\beta_{t-1}}\,\bar{\varepsilon}_{t-2}\right) + \sqrt{\beta_t}\,\bar{\varepsilon}_{t-1} =$$

$$= \sqrt{\alpha_t\alpha_{t-1}}\cdot\bar{x}_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\cdot\bar{\varepsilon}_{t-2} + \sqrt{1-\alpha_t}\,\bar{\varepsilon}_{t-1} =$$

$$\mathcal{N}(\bar{0}, \alpha_t(1-\alpha_{t-1})I) + \mathcal{N}(\bar{0}, (1-\alpha_t)I)$$

$$= \sqrt{\alpha_t\alpha_{t-1}}\,\bar{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\cdot\bar{\varepsilon} = \dots$$

$\alpha_1 \dots \alpha_t = A_t$

$$\bar{x}_t = \sqrt{\alpha_t\alpha_{t-1}\dots\alpha_1}\,\bar{x}_0 + \sqrt{1-\alpha_t\alpha_{t-1}\dots\alpha_1}\cdot\bar{\varepsilon}$$

$$q(\bar{x}_t|\bar{x}_0) = \mathcal{N}(\sqrt{A_t}\cdot\bar{x}_0, (1-A_t)\cdot I)$$

$$q(\bar{x}_{t-1}|\bar{x}_t, \bar{x}_0) = \frac{q(\bar{x}_{t-1}|\bar{x}_0)\,q(\bar{x}_t|\bar{x}_{t-1}, \bar{x}_0)}{q(\bar{x}_t|\bar{x}_0)} =$$

$\to A_{t-1}$ $\quad \to \sqrt{1-\beta_t}\dots$

$$= \frac{\mathcal{N}(\bar{x}_{t-1}|\sqrt{A_{t-1}}\cdot\bar{x}_0, (1-A_{t-1})I)\cdot\mathcal{N}(\bar{x}_t|\sqrt{\alpha_t}\,\bar{x}_{t-1}, \beta_t I)}{\mathcal{N}(\bar{x}_t|\sqrt{A_t}\,\bar{x}_0, (1-A_t)I)} =$$

$$= \text{Const} \cdot e^{-\frac{1}{2}\left(\frac{(\bar{x}_{t-1} - \sqrt{A_{t-1}}\,\bar{x}_0)^2}{1-A_{t-1}} + \frac{(\bar{x}_t - \sqrt{\alpha_t}\,\bar{x}_{t-1})^2}{\beta_t} - \cancel{\frac{(\bar{x}_t - \sqrt{A_t}\,\bar{x}_0)^2}{1-A_t}} \right)}$$

$$\bar{x}_{t-1}^T \bar{x}_{t-1} : \quad \frac{1}{1-A_{t-1}} + \frac{\alpha_t}{\beta_t}$$

$$\bar{x}_{t-1} : \quad -\frac{2\sqrt{A_{t-1}}}{1-A_{t-1}} \cdot \bar{x}_0 - \frac{2\sqrt{\alpha_t}}{\beta_t}\,\bar{x}_t$$

$$= \frac{1}{\tilde{\beta}_t}\left(\bar{x}_{t-1} - \tilde{\mu}(\bar{x}_t, \bar{x}_0)\right)^2 + \text{const}$$

$$\tilde{\beta}_t = \frac{1}{\frac{1}{1-A_{t-1}} + \frac{\alpha_t}{\beta_t}} = \frac{\beta_t(1-A_{t-1})}{\underbrace{\beta_t + \alpha_t}_{=1} \underbrace{- \alpha_t A_{t-1}}_{A_t}} = \beta_t \cdot \frac{1-A_{t-1}}{1-A_t}$$

$$\bar{x}_t = \sqrt{A_t} \cdot \bar{x}_0 + \sqrt{1-A_t} \cdot \bar{\varepsilon}$$

$$\tilde{\mu}(\bar{x}_t, \bar{x}_0) = \sqrt{\alpha_t} \cdot \frac{1-A_{t-1}}{1-A_t} \cdot \bar{x}_t + \frac{\beta_t \sqrt{A_{t-1}}}{1-A_t}\,\bar{x}_0 =$$

$$= \sqrt{\alpha_t} \cdot \frac{1-A_{t-1}}{1-A_t}\,\bar{x}_t + \frac{\beta_t \cancel{\sqrt{A_{t-1}}}}{1-A_t} \cdot \frac{\bar{x}_t - \sqrt{1-A_t} \cdot \bar{\varepsilon}}{\cancel{\sqrt{A_t}}\ \sqrt{\alpha_t}} =$$

$$\alpha_t - A_t + 1 - \alpha_t = \cancel{1-A_t}$$

$$= \frac{\alpha_t(1-A_{t-1}) + \beta_t}{\cancel{(1-A_t)}\sqrt{\alpha_t}} \cdot \bar{x}_t - \frac{\beta_t}{\sqrt{\alpha_t(1-A_t)}}\,\bar{\varepsilon}$$

$$q(\bar{x}_{t-1} | \bar{x}_t, \bar{x}_0) = \mathcal{N}\left(\bar{x}_{t-1} \,\middle|\, \frac{1}{\sqrt{\alpha_t}} \cdot \bar{x}_t - \frac{\beta_t}{\sqrt{\alpha_t(1-A_t)}}\,\bar{\varepsilon},\ \beta_t \cdot \frac{1-A_{t-1}}{1-A_t} \cdot I\right)$$

$$\underline{p_\theta(\bar{x}_0)} \approx q(\bar{x}_0)$$

$$KL(q(\bar{x}_0) \,\|\, p_\theta(\bar{x}_0)) \to \min$$

$$\int q \log \frac{q}{p_\theta}\, d\hat{x}_0 = \int q \log q \, d\hat{x}_0 - \boxed{\int q(\bar{x}_0) \log p_\theta(\bar{x}_0)\, d\bar{x}_0} \to \max$$

$$E_{q(\bar{x}_0)}\left[\log p_\theta(\bar{x}_0)\right] = E_{q(\bar{x}_0)}\left[\log \int p_\theta(\bar{x}_0, \bar{x}_1, \ldots \bar{x}_T)\, d\bar{x}_1 \ldots d\bar{x}_T\right]$$

$$E_{q(\bar{x}_1 \ldots \bar{x}_T | \bar{x}_0)}\left[\div\right]$$

$$= E_{q(\bar{x}_0)}\left[\log \overline{\int q(\bar{x}_1 \ldots \bar{x}_T | \bar{x}_0) \cdot \frac{p_\theta(\bar{x}_{0 \ldots T})}{q(\bar{x}_{1 \ldots T} | \bar{x}_0)}\, d\bar{x}_1 \ldots d\bar{x}_T}\right] \geq$$

$$\geq E_{q(\bar{x}_0)}\left[E_{q(\bar{x}_1 \ldots \bar{x}_T | \bar{x}_0)}\left[\log \frac{p_\theta(\bar{x}_{0 \ldots T})}{q(\bar{x}_{1 \ldots T} | \bar{x}_0)}\right]\right] \to \max$$

$$E_{q(\bar{x}_0, \bar{x}_1, \ldots \bar{x}_T)}\left[\log \frac{q(\bar{x}_1 \ldots \bar{x}_T | \bar{x}_0)}{\underline{p_\theta(\bar{x}_0, \bar{x}_1 \ldots \bar{x}_T)}}\right] \to \min$$

$$\underline{p_\theta(\bar{x}_0, \bar{x}_1 \ldots \bar{x}_T)} = p_\theta(\bar{x}_T)\, p_\theta(\bar{x}_{T-1} | \bar{x}_T) \ldots p_\theta(\bar{x}_0 | \bar{x}_1)$$

$$p_\theta(\bar{x}_{T-1} | \bar{x}_T) = \mathcal{N}(\bar{x}_{T-1} | \bar{\mu}_\theta(\bar{x}_T), \sigma_T^2)$$

$$\|$$

$$E_q\left[\log \frac{q(\bar{x}_1 | \bar{x}_0)\, q(\bar{x}_2 | \bar{x}_1, \bar{x}_0) \ldots q(\bar{x}_T | \bar{x}_{T-1}, \bar{x}_0)}{p_\theta(\bar{x}_T)\, p_\theta(\bar{x}_{T-1} | \bar{x}_T) \ldots p_\theta(\bar{x}_0 | \bar{x}_1)}\right] =$$

$$= E_q\left[-\log p_\theta(\bar{x}_T) + \sum_{t=2}^{T} \log \frac{q(\bar{x}_t | \bar{x}_{t-1}, \bar{x}_0)}{p_\theta(\bar{x}_{t-1} | \bar{x}_t)} + \log \frac{q(\bar{x}_1 | \bar{x}_0)}{p_\theta(\bar{x}_0 | \bar{x}_1)}\right] =$$

$$\sum_{t=2}^{T} \log \frac{q(\bar{x}_{t-1} | \bar{x}_t, \bar{x}_0)\, \boxed{q(\bar{x}_t | \bar{x}_0)}}{p_\theta(\bar{x}_{t-1} | \bar{x}_t) \cdot \boxed{q(\bar{x}_{t-1} | \bar{x}_0)}}$$

$$= \mathbb{E}_q\left[-\log p_\theta(\bar{x}_T) + \sum_{t=2}^{T} \log \frac{q(\bar{x}_{t-1}|\bar{x}_t,\bar{x}_0)}{p_\theta(\bar{x}_{t-1}|\bar{x}_t)} + \log \frac{q(\bar{x}_T|\bar{x}_0)}{\cancel{q(\bar{x}_1|\bar{x}_0)}} \cdot \frac{\cancel{q(\bar{x}_1|\bar{x}_0)}}{p_\theta(\bar{x}_0|\bar{x}_1)}\right]$$

$$= \mathbb{E}_q\left[\underbrace{\log \frac{q(\bar{x}_T|\bar{x}_0)}{p_\theta(\bar{x}_T)}}_{L_T} + \underbrace{\sum_{t=2}^{T} \log \frac{q(\bar{x}_{t-1}|\bar{x}_t,\bar{x}_0)}{p_\theta(\bar{x}_{t-1}|\bar{x}_t)}}_{L_{t-1}} \underbrace{- \log p_\theta(\bar{x}_0|\bar{x}_1)}_{L_0}\right]$$

$$\beta_{t+1} \frac{1-A_t}{1-A_{t+1}} I$$

$L = L_0 + L_1 + \dots + L_{T-1} + L_T$, ye $\mathcal{N}(\bar{x}_t | \overbrace{\tilde{\mu}_\theta(\bar{x}_{t+1}, t)}^{\sim}, \overbrace{\Sigma_\theta(\bar{x}_{t+1}, t)}^{\sim})$

$L_0 = -\log p_\theta(\bar{x}_0|\bar{x}_1)$,

$L_t = KL(q(\bar{x}_t|\bar{x}_{t+1},\bar{x}_0) \,\|\, p_\theta(\bar{x}_t|\bar{x}_{t+1}))$

$L_T = KL(q(\bar{x}_T|\bar{x}_0) \,\|\, p_\theta(\bar{x}_T))$