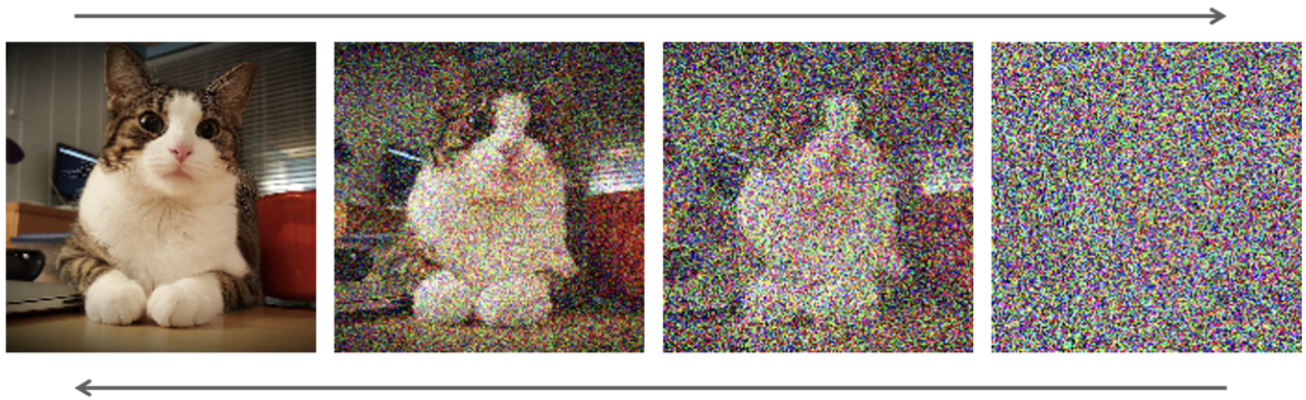
TTIC 31230, Fundamentals of Deep Learning

David McAllester, Autumn 2022

Markovian VAEs

Markovian VAEs

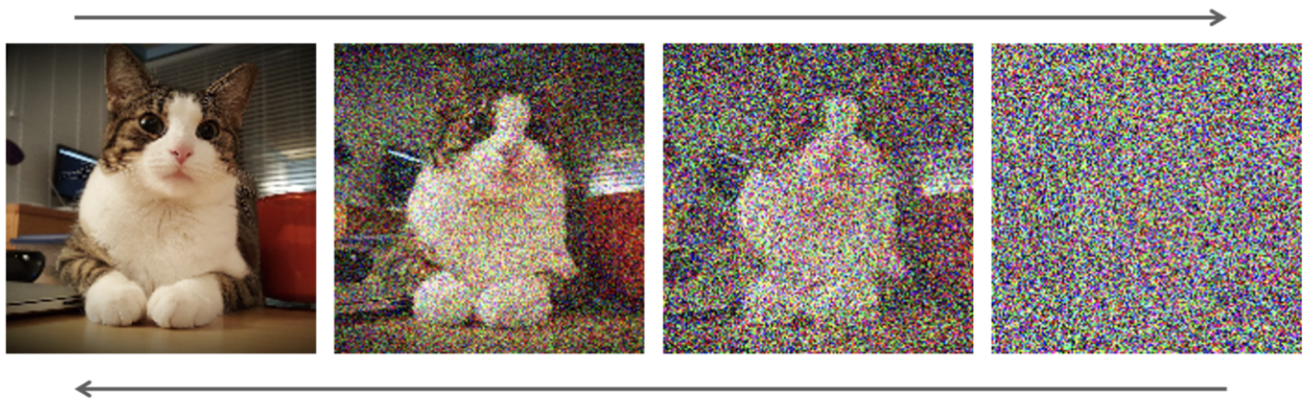
A diffusion models computes and inverts a sequence



So does an autoregressive language model

[Sally talked to John] $\rightleftarrows$ [Sally talked to] $\rightleftarrows$ [Sally talked] $\rightleftarrows$ [Sally]

Markovian VAEs



[Sally talked to John] $\rightleftarrows$ [Sally talked to] $\rightleftarrows$ [Sally talked] $\rightleftarrows$ [Sally]

$$z_0 \rightleftarrows z_1 \rightleftarrows \cdots \rightleftarrows z_L$$

Markovian VAEs

$$z_0 \overset{\rightarrow}{\leftarrow} z_1 \overset{\rightarrow}{\leftarrow} \cdots \overset{\rightarrow}{\leftarrow} z_L$$

Encoder: $\mathrm{Pop}(z_0)$ and $P_{\mathrm{enc}}(z_{\ell+1}|z_\ell)$.

Generator: $P_{\mathrm{pri}}(z_L)\ P_{\mathrm{gen}}(z_{\ell-1}|z_\ell)$.

The encoder and the decoder define distributions $P_{\mathrm{enc}}(z_0, \ldots, z_L)$ and $P_{\mathrm{gen}}(z_0, \ldots, z_N)$ respectively.

VAE Review

A variational autoencoder (VAE) has only z_0 (previously written y) and z_1 (previously written z).

$$P_{\mathrm{enc}}(z_0, z_1) \;=\; \mathrm{Pop}(z_0) P_{\mathrm{enc}}(z_1|z_0)$$

$$P_{\mathrm{gen}}(z_0, z_1) \;=\; P_{\mathrm{pri}}(z_1) P_{\mathrm{gen}}(z_0|z_1)$$

The Single Layer ELBO

$$H(z_0) = E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{enc}}(z_0)P_{\mathrm{enc}}(z_1|z_0)}{P_{\mathrm{enc}}(z_1|z_0)}\right]$$

$$= E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{enc}}(z_1)P_{\mathrm{enc}}(z_0|z_1)}{P_{\mathrm{enc}}(z_1|z_0)}\right]$$

$$\leq E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{gen}}(z_1)P_{\mathrm{gen}}(z_0|z_1)}{P_{\mathrm{enc}}(z_1|z_0)}\right] \quad \text{cross-entropy bounds entropy}$$

$$= E_{\mathrm{enc}}\ KL(P_{\mathrm{enc}}(z_1|z_0),\ P_{\mathrm{gen}}(z_1)) + E_{\mathrm{enc}}[-\ln P_{\mathrm{gen}}(z_0|z_1)]$$

The Markovian ELBO

$$
\begin{aligned}
H(z_0) &= E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{enc}}(z_0)P_{\mathrm{enc}}(z_1,\ldots,z_L|z_0)}{P_{\mathrm{enc}}(z_1,\ldots,z_L|z_0)}\right] \\
&= E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{enc}}(z_0|z_1)P_{\mathrm{enc}}(z_1|z_2)\cdots P_{\mathrm{enc}}(z_{L-1}|z_L)P_{\mathrm{enc}}(z_L)}{P_{\mathrm{enc}}(z_1|z_2,z_0)\cdots P_{\mathrm{enc}}(z_{L-1}|z_L,z_0)P_{\mathrm{enc}}(z_L|z_0)}\right] \\
&\leq E_{\mathrm{enc}}\left[-\ln \frac{P_{\mathrm{gen}}(z_0|z_1)P_{\mathrm{gen}}(z_1|z_2)\cdots P_{\mathrm{gen}}(z_{L-1}|z_L)P_{\mathrm{gen}}(z_L)}{P_{\mathrm{enc}}(z_1|z_2,z_0)\cdots P_{\mathrm{enc}}(z_{L-1}|z_L,z_0)P_{\mathrm{enc}}(z_L|z_0)}\right] \\
&= \left\{\begin{array}{l}
E_{\mathrm{enc}}\left[-\ln P_{\mathrm{gen}}(z_0|z_1)\right] \\
\\
+\sum_{i=2}^{L}\; E_{\mathrm{enc}}\; KL(P_{\mathrm{enc}}(z_{i-1}|z_i,z_0),\;\; P_{\mathrm{gen}}(z_{i-1}|z_i)) \\
\\
+E_{\mathrm{enc}}\; KL(P_{\mathrm{enc}}(Z_L|z_0), p_{\mathrm{gen}}(Z_L))
\end{array}\right.
\end{aligned}
$$

END