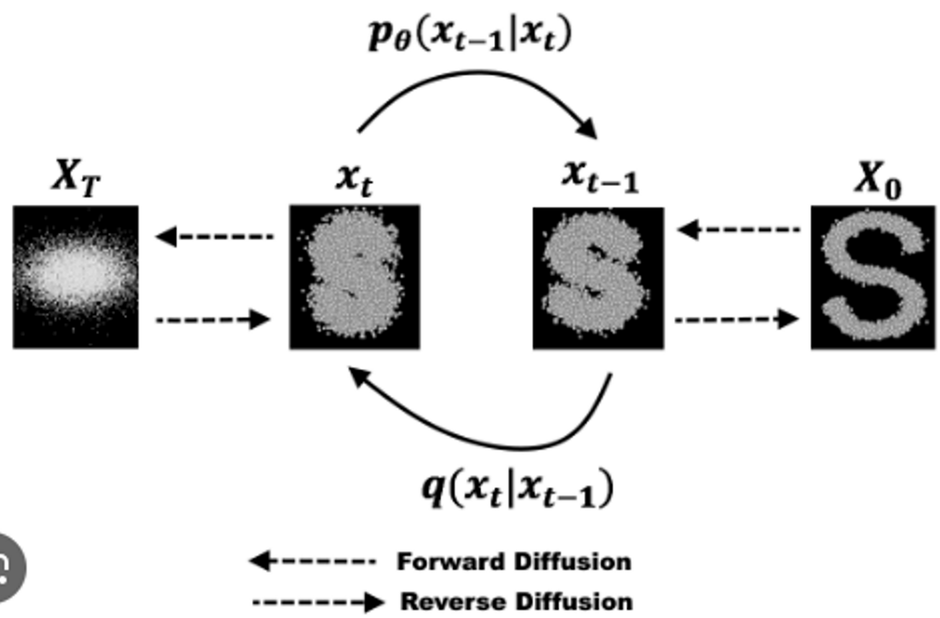
Project PT

17 February 2024 21:16

Explaination of Denoising Diffusion Probabilistic Models(DDPM)

Defusion Models:

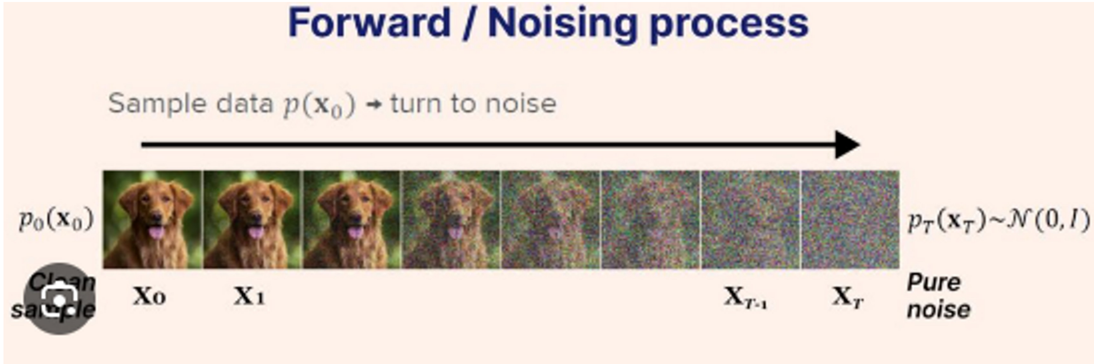
The essential idea is to systematically and slowly destroy structure in a data distribution through an iterative forward deffusion process . We then learn a reverse deffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data



Key Concepts of Deffusion Models

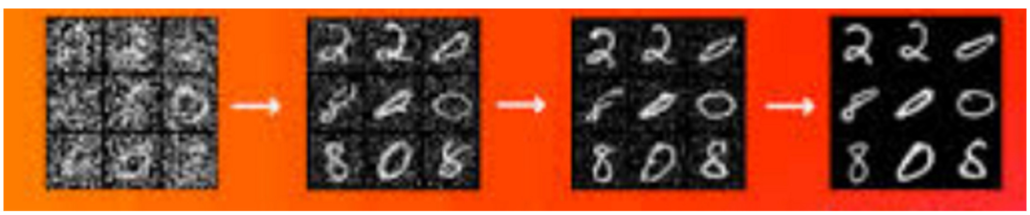
1. Forward Deffusion Process:

   The forward process is a Markov chain that gradually adds noise to the data, effectively transforming a data distribution into a simple noise distribution (usually Gaussian). This process is parameterized by a time-dependent noise schedule.



2. Reverse Deffusion Process:

The reverse process is the generative aspect of the model, where the goal is to start from pure noise and gradually denoise it back into a data sample.
This reverse process is also modeled as a Markov chain but in the opposite direction of the forward process. It involves learning a series of denoising steps that approximate the reverse of the noise addition in the forward process.



-

Mathematical Intuition :-

Date ___ Page ___

Notation :

x_t = (image at time stamp t)

x_0 = (Denote original image)

x_{42} = (Denote noisy image after 42 iteration of noise)

Forward process

$$q(x_t \mid x_{t-1})$$

q ... more noisy image; x_{t-1} ... Less noisy image

$$q(x_t \mid x_{t-1}) = N(x_t, \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \quad \text{--- (1)}$$

N: Normal Distribution; x_t: output; $\sqrt{1-\beta_t}\, x_{t-1}$: mean; $\beta_t I$: variance

Reverse process

$$p(x_{t-1} \mid x_t)$$

$$\alpha_t = 1 - \beta_t \quad \text{--- (2)}$$

cummalative product of all alphas from 0 to t

$$\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s \quad \text{--- (3)}$$

eg: $t = 4$

$$\bar{\alpha}_4 = \alpha_1 . \alpha_2 . \alpha_3 . \alpha_4$$

From eq^n (1)

$$q(x_t \mid x_{t-1}) = N(x_t, \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I)$$

from Reparametrized trick

$$N(\mu, \sigma^2) = \mu + \sigma \cdot \varepsilon \quad ; \quad \varepsilon \sim N(0,1)$$

$$\therefore \quad q(x_t \mid x_{t-1}) = \sqrt{(1-\beta_t)}\, x_{t-1} + \sqrt{\beta_t} \cdot \varepsilon$$

Date / /
Page

$$q(x_t|x_{t-1}) = \sqrt{\alpha_t}\; x_{t-1} + \sqrt{(1-\alpha_t)} \cdot \varepsilon$$

$$= \sqrt{\alpha_t \cdot \alpha_{t-1}}\; x_{t-2} + \sqrt{(1-\alpha_t \cdot \alpha_{t-1})} \cdot \varepsilon$$

$$= \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2}}\; x_{t-3} + \sqrt{1-\alpha_t \cdot \alpha_{t-1} \alpha_{t-2}} \cdot \varepsilon$$

$$= \ldots$$

$$q(x_t|x_{t-1}) = \sqrt{\alpha_t \alpha_{t-1} \alpha_{t-2} \ldots \alpha_1}\; x_0 + \sqrt{1-\alpha_t \alpha_{t-1} \ldots \alpha_1} \cdot \varepsilon$$

$$q(x_t|x_{t-1}) = \sqrt{\bar{\alpha}_t}\; x_0 + \sqrt{(1-\bar{\alpha}_t)}\; \varepsilon \quad \text{—— (4)} \quad \{\text{from (3)}\}$$

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{\bar{\alpha}_t}\, x_0; (1-\bar{\alpha}_t) I) \quad \text{—— (5)}$$

and

$$q(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

Loss function

$$-\log(P_\theta(x_0)) = \text{negative log likelihood}$$

The probability of x_0 is nicely computible as a depends on all other times steps coming before x_0 starting the x_t this with me keeping a track of $t-1$ on other random variables which is just a possible practice as a solution we can compute variational lower bound for this objective.

$$-\log(P_\theta(x_0)) \le -\log(P_\theta(x_0)) + D_{KL}(q(x_{1:T}|x_1) \,||\, P_\theta(x_{1:T}|x_0)) \quad \text{—— (6)}$$

Deffusion models ~ Variational Autoencoder

$$D_{KL}(P||q) = \int_x p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

Date / /
Page

$$= -\log(p(x_T)) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right) + \log\left(\frac{q(x_T|x_0)}{q(x_1|x_0)}\right) + \log\left(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right)$$

last two term

$$\log\left(\frac{q(x_T|x_0)}{q(x_1|x_0)}\right) + \log\left(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right) = \log(q(x_T|x_0)) - \log(q(x_1|q(x_0))) + \log(q(x_1|q(x_0))) - \log(p_\theta(x_0|x_1))$$

$\therefore$

$$= -\log(p(x_T)) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right) + \log(q(x_T|x_0)) - \log(p_\theta(x_0|x_1))$$

$$= -\log(p(x_T)) + \log(q(x_T|x_0)) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right) - \log(p_\theta(x_0|x_1))$$

$$= \log\left(\frac{q(x_T|x_0)}{p(x_T)}\right) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right) - \log(p_\theta(x_0|x_1))$$

$$= D_{KL}(q(x_T|x_0) \| p(x_T)) + \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)) - \log(p_\theta(x_0|x_1))$$

Ignoring first ~~con~~ term completly

$$= \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)) - \log(p_\theta(x_0|x_1)) \quad \text{--- (8)}$$

$$\mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \qquad \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

$$\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta I)$$

Now

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{(\alpha_t)}(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)} x_t + \frac{\sqrt{(\bar{\alpha}_{t-1})}\,\beta_t}{(1-\bar{\alpha}_t)} x_0$$

$$\tilde{\beta}_t = \frac{(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)} \cdot \beta_t$$

Date / /
Page

$$= -\log(P(x_T)) + \sum_{t=1}^{T} \log\left(\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right)$$

$$\log\left(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}\right) = -\log(P(x_T)) + \log\left(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right) + \sum_{t=2}^{T} \log\left(\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}\right)$$

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)\,q(x_t)}{q(x_{t-1})} \quad \text{(Bayes formula)}$$

$$\Rightarrow = \frac{q(x_{t-1}|x_t, x_0)\,q(x_t|x_0)}{q(x_{t-1}|x_0)}$$

$$= -\log(P(x_T)) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)\,q(x_t|x_0)}{p_\theta(x_{t-1}|x_t)\,q(x_{t-1}|x_0)}\right) + \log\left(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right).$$

$$q(x_1|x_0) = \frac{q(x_0|x_1)\,q(x_1)}{q(x_0)} = \frac{q(x_0|x_1, x_0)\,q(x_1|x_0)}{q(x_0|x_0)}$$

$$= -\log(P(x_T)) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\right) + \sum_{t=2}^{T} \log\left(\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}\right) + \log\left(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)}\right)$$

for second summation

let $T = 4$

$$\sum_{t=2}^{4} \log\left(\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}\right) = \log\left(\prod_{t=2}^{4} \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}\right) = \log\left(\frac{q(x_2|x_0)\,q(x_3|x_0)\,q(x_4|x_0)}{q(x_1|x_0)\,q(x_2|x_0)\,q(x_3|x_0)}\right)$$

$$= \log\left(\frac{q(x_4|x_0)}{q(x_1|x_0)}\right)$$

$\therefore$ In general

$$\sum_{t=2}^{T} \log\left(\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}\right) = \log\left(\frac{q(x_T|x_0)}{q(x_1|x_0)}\right)$$

Date ..../..../....
Page........

$$\log\left(\frac{q(x_{1}:T|x_0)}{P_\theta(x_1:T|x_0)}\right)$$

$$P_\theta(x_1:T|x_0) = \frac{P_\theta(x_0|x_1:T)\cdot P_\theta(x_1:T)}{P_\theta(x_0)}$$ { Bayes formula

$\dfrac{P_\theta(x_0|x_1:T)}{P_\theta(x_0)}$ joint probability

$$\frac{P_\theta(x_0:T)}{P_\theta(x_0)}$$

$$\therefore \log\left(\frac{q(x_1:T|x_0)}{P_\theta(x_1:T|x_0)}\right) = \log\left(\frac{q(x_1:T|x_0)}{\frac{P_\theta(x_0:T)}{P_\theta(x_0)}}\right)$$

$$\log\frac{q(x_1:T|x_0)}{P_\theta(x_1:T|x_0)} = \log\left(\frac{q(x_1:T|x_0)}{P_\theta(x_0:T)}\right) + \log(P_\theta(x_0))$$

$\therefore$ from eq^n (6)

$$-\log(P_\theta(x_0)) \le -\log(P_\theta(x_0)) + \log\left(\frac{q(x_1:T|x_0)}{P_\theta(x_0:T)}\right) + \log(P_\theta(x_0))$$

$$\Rightarrow -\log(P_\theta(x_0)) \le \log\left(\frac{q(x_1:T|x_0)}{P_\theta(x_0:T)}\right) \quad (7)$$

Which is variational lower bound.

where

$$P_\theta(x_0:T) = P(T)\prod_{t=1}^{T} P_\theta(x_{t-1}|x_t)$$ (from Bayes theo.)

Now $$\log\left(\frac{q(x_1:T|x_0)}{P_\theta(x_0:T)}\right) = \log\left(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{P(x_T)\prod_{t=1}^{T} P_\theta(x_{t-1}|x_t)}\right)$$

$$= -\log(P(x_T)) + \log\left(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} P_\theta(x_{t-1}|x_t)}\right)$$

Date
Page

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)}x_t + \frac{\sqrt{(\bar{\alpha}_{t-1})}\,\beta_t}{(1-\bar{\alpha}_t)}x_0$$

from eq^n (4)

$$x_t = \sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{(1-\bar{\alpha}_t)}\,\varepsilon$$

$$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{(1-\bar{\alpha}_t)}\,\varepsilon\right)$$

$$\therefore \tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\,\beta_t}{(1-\bar{\alpha}_t)}\cdot\frac{1}{\sqrt{\bar{\alpha}_t}}\left(x_t - \sqrt{1-\bar{\alpha}_t}\,\varepsilon\right)$$

$$\boxed{\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon\right)}$$

from eq^n (8)

$$\therefore \log\left(\frac{q(x_{1:T}\mid x_0)}{p_\theta(x_{0:T})}\right) = \sum_{t=2}^{T} D_{KL}\left(q(x_{t-1}\mid x_t, x_0)\,\|\,p_\theta(x_{t-1}\mid x_t)\right) - \log\left(p_\theta(x_0\mid x_1)\right)$$

$$= N(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)$$

$$= N(x_{t-1}; \mu_\theta(x_t, t), \beta I)$$

$$\tilde{\mu}_t = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon\right)$$

Mean Squared Error (Actual μ - Predicted μ)

$$L_t = \frac{1}{2\sigma_t^2}\left\|\left(\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\right)\right\|^2$$

$$L_t = \frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{(1-\bar{\alpha}_t)}}\,\varepsilon\right) - \mu_\theta(x_t, t)\right\|^2$$

Date / /
Page

$$\because \mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon_\theta(x_t, t)\right)$$

$$L_t = \frac{1}{2\sigma_t^2}\left\|\left(\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon_\theta(x_t, t)\right)\right)\right\|^2$$

$$L_t = \frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\hat{\alpha}_t)}\left\|\left(\varepsilon - \varepsilon_\theta(x_t, t)\right)\right\|^2$$

$$\boxed{L_t = \left\|\left(\varepsilon - \varepsilon_\theta(x_t, t)\right)\right\|^2}$$

$$L_t = \left\|\varepsilon - \varepsilon_\theta\left(\sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\varepsilon, t\right)\right\|^2$$

$$\downarrow$$

$$N\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right)$$

$$N\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon_\theta(x_t, t)\right), \beta_t\right)$$

Parameterization

$$N(\mu, \sigma^2) = \mu + \sigma\varepsilon$$

$$\therefore \boxed{\underbrace{x_{t-1}}_{\text{less noisy}} = \frac{1}{\sqrt{\alpha_t}}\underbrace{\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\varepsilon_\theta(x_t, t)\right)}_{\text{more noisy}} + \sqrt{\beta_t}\,\varepsilon}$$

$$L_{VLB} = \underbrace{\sum_{t=2}^{T} D_{KL}\left(q(x_{t-1}|x_t, x_0) \,\|\, p_\theta(x_{t-1}|x_t)\right)}_{\|\epsilon - \epsilon_\theta(x_t, t)\|^2} - \log\left(p_\theta(x_0|x_1)\right)$$

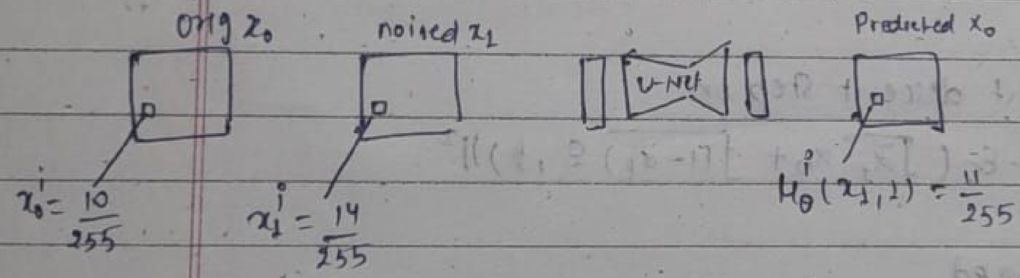
Dimension of data

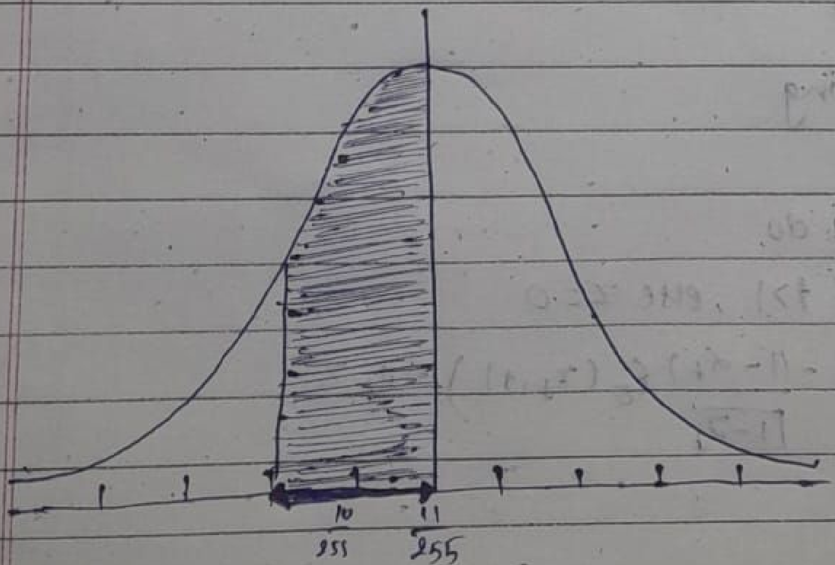
$$p_\theta(x_0|x_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}\left(x; \mu_\theta^i(x_1, 1), \beta_1\right) dx$$

$$\delta_+(x) = \begin{cases} \infty & \text{if } x = 1 \\ x + \frac{1}{255} & \text{if } x < 1 \end{cases}$$

$$\delta_-(x) = \begin{cases} -\infty & \text{if } x = -1 \\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$$

These just give the border for the integral and say that we integrate from one pixel below to one pixel above





Date
Page

$t > 1$

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(x_t,t)\right) + \sqrt{\beta_t}\,\varepsilon$$

$t = 1$:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(x_t,t)\right)$$

$$L_{simple} = \mathbb{E}_{t,x_0,\varepsilon}\left[\| \varepsilon - \varepsilon_\theta(\underbrace{\sqrt{(\bar{\alpha}_t)}\, x_0 + \sqrt{(1-\bar{\alpha}_t)}\,\varepsilon}_{x_t}, t)\|^2\right]$$

$$\boxed{L_{simple} = \mathbb{E}_{t,x_0,\varepsilon}\left[\| \varepsilon - \varepsilon_\theta(x_t, t)\|^2\right]}$$

Algorithm 1 Trainning

1. Repeat
2. $x_0 \sim q(x_0)$
3. $t \sim \text{Uniform}(\{1, \dots, T\})$
4. $\varepsilon \sim N(0, I)$,
5. Take gradient descent step on
   $\nabla_\theta \| \varepsilon - \varepsilon_\theta(\sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{(1-\bar{\alpha}_t)}\,\varepsilon, t)\|^2$
6. untill converged.

Algorithm 2 Sampling

1. $x_T \sim N(0, I)$
2. for $t = T, \dots, 1$ do
3. $z \sim N(0, I)$ if $t > 1$, else $z = 0$
4. $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{(1-\alpha_t)}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(x_t, t)\right) + \sigma_t z$.
5. end for
6. return x_0