FORWARD DIFFUSION PROCESS

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

Gaussian Distribution — Mean — Hyperparam. — Covariance matrix

Key Points

- True Markov Process: x_t only depends on x_{t-1}
- I is the identity matrix, indicating every dimension receives independent noise w/ variance β_t.
- $\beta_t = 0 \Rightarrow$ no noise, full signal
- $\beta_t = 1 \Rightarrow$ all noise, no signal

$x_0 \to x_1 \to x_2 \to \dots \to x_T$

$$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

Joint Distribution over entire trajectory.

we can do this thanks to the Markov property.

Key Points

- Variance Scheduling: how β_t changes over time
- Total variance of x_t comes from two sources:
  1 - Scaled variance of the previous state (μ_t term)
  2 - The variance of the newly added noise (Σ_t term)
- Why $\sqrt{1-\beta_t}$? To keep variance stable and to destroy original image x_0 in a controlled manner.

Stable variance

$$var(aX + bY) = a^2 var(X) + b^2 var(Y)$$

If $Var(x_0) = 1$, then

$$var(x_1) = \sqrt{1-\beta_1}^{\,2}\ var(x_0) + \beta_1\ var(\varepsilon_t)$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ — unit variance — 1 by definition

$$var(x_1) = (1-\beta_1)(var(x_0)) + \beta_1\ Var(\varepsilon_t)$$