Variational AutoEncoder

Hiroki Adachi

Source code: https://github.com/hirokiadachi/VAE

MACHINE PERCEPTION AND ROBOTICS GROUP

生成モデルとは

→ **CNN**
(Convolutional Neural Networks) → Cat

$\mathbf{z}$
(Latent vector) → **CNN**
(Convolutional Neural Networks) →

どうやって獲得する？

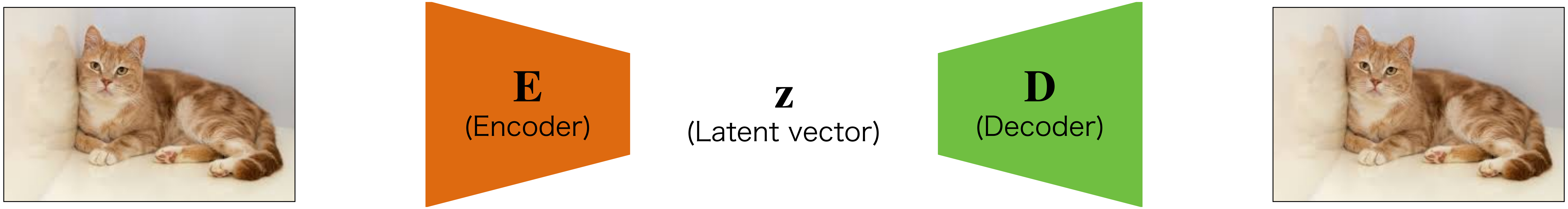
Autoencoder (AE)

MPRG

- EncoderとDecoderによって構成
  - Encoder：入力データの情報を低次元ベクトルへ埋め込む
  - Decoder：低次元ベクトルを入力としてデータを復元

- 入力データを低次元ベクトルから再構成できるように学習
  - データを圧縮または復元するために適した重みパラメータを獲得

- 深層学習をベースとした識別や認識などのfine-tuningとして利用可能



AEの派生手法

- Sparse AE
  - 活性化平均度によって潜在変数に関してスパース性能が改善

    → High sparseness：再構成データがシンプルな表現になり，数字がそのまま出力

    → Low sparseness：潜在変数が様々な情報を含むため，複雑なデータが復元される

- Denoising AE
  - 欠落したデータを入力して，補間するように学習することが可能
    - 入力データはガウスノイズ，胡麻塩ノイズ，マスクなどの処理を施す

- Variational AE
  - 変分推論を使用
    - 中間層で獲得する潜在変数が何かしらの確率密度関数（PDF）に基づくように学習

AEの派生手法

MPRG

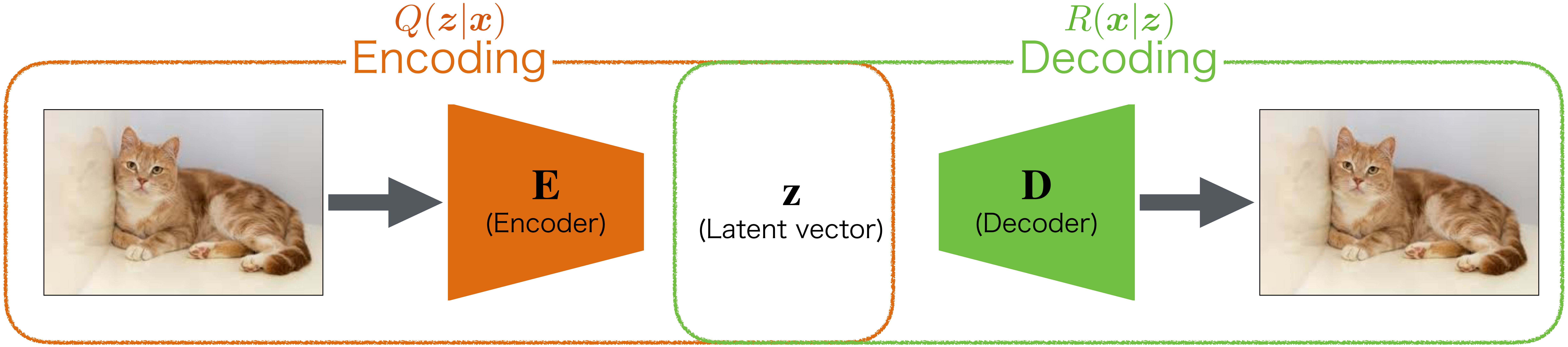
- Sparse AE
  - 活性化平均度によって潜在変数に関してスパース性能が改善
    → High sparseness：再構成データがシンプルな表現になる
    → Low sparseness：潜在変数が様々な情報を含むため，複雑なデータが復元される

- Denoising AE
  - 欠落したデータを入力して，保管するように学習することが可能
    - 入力データはガウスノイズ，胡麻塩ノイズ，マスクなどの処理を施す

- **Variational AE**
  - **変分推論を使用**
    - **中間層で獲得する潜在変数が何かしらの確率密度関数（PDF）に基づくように学習**

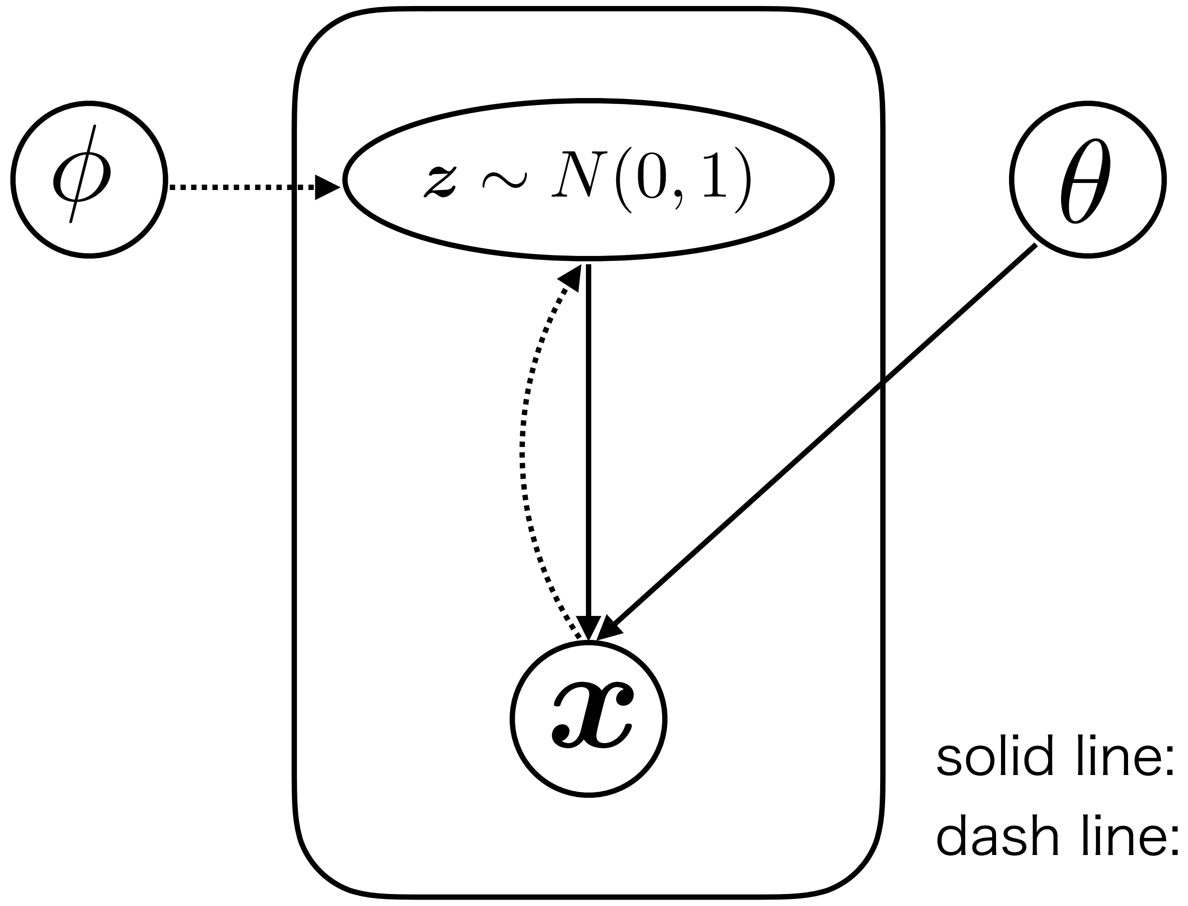
Variational Autoencdoer (VAE)

- 生成モデルという枠組みでAEを解釈した手法
  - ネットワークの構造はAEと似ている



<u>**VAE motivation: Maximising**</u>

$$p(\boldsymbol{x};\theta) = \int \underbrace{p(\boldsymbol{x}|\boldsymbol{z};\theta)}_{\text{posterior}} \underbrace{p(\boldsymbol{z};\theta)}_{\text{prior}} d\boldsymbol{z}$$



solid line: decoder
dash line: encoder

解決策

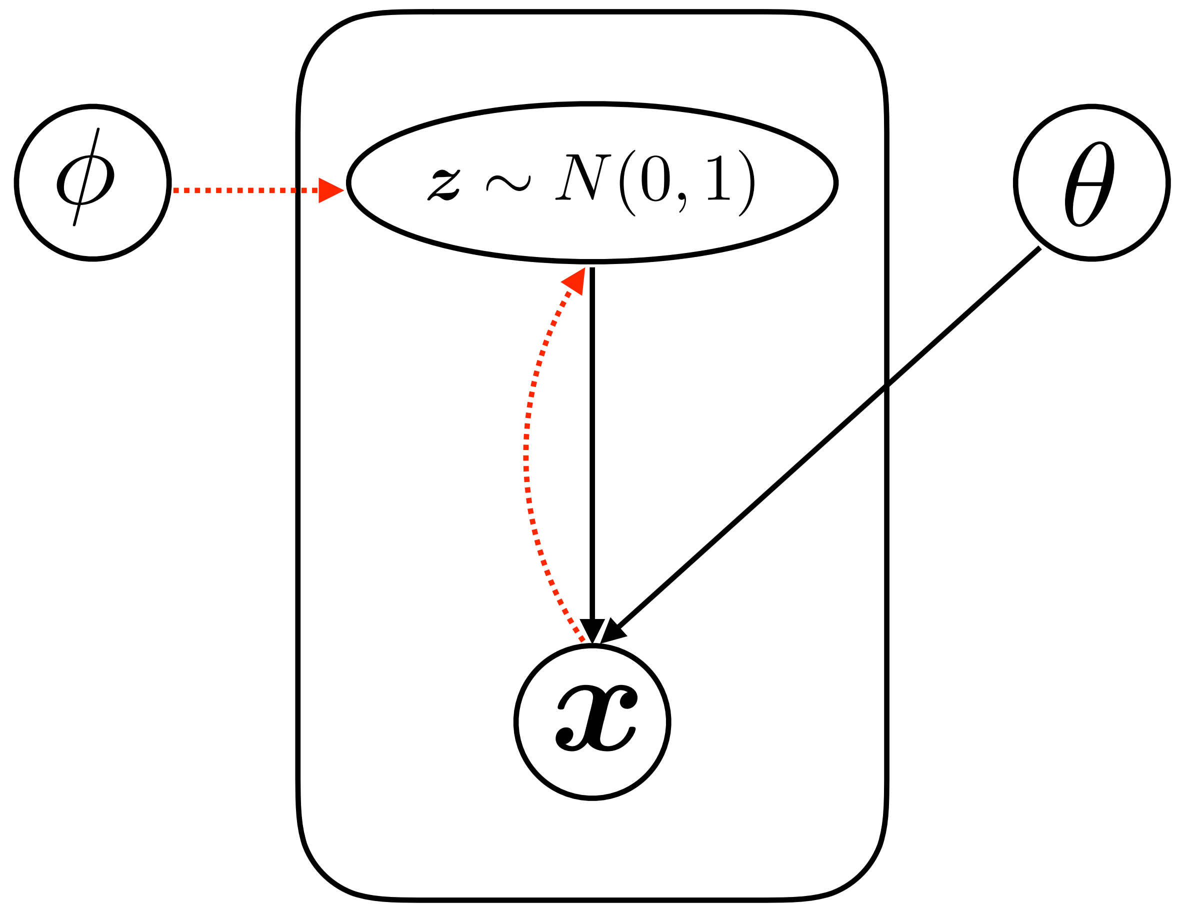
- 古典的または一般的
  - EM algorithm，MCMC (Marcov Chain Monte Carlo)などを使用

- VAE
  - 従来とは異なる変分推論手法を使用
    - Evidence Lower Bound (ELBO)

Evidence Lower Bound

$$p(\boldsymbol{x};\theta) = \int \underline{p(\boldsymbol{x}|\boldsymbol{z};\theta)}p(\boldsymbol{z};\theta)d\boldsymbol{z}$$

$$q(\boldsymbol{z}|\boldsymbol{x};\phi)$$



p: decoder
q: encoder

- p(**z**|**x**; θ)は直接推定することは困難
- 推定が困難なp(**z**|**x**; θ)の代わりにq(**z**|**x**; ϕ)を使用
- KLを利用して分布の類似度を調査：$\boxed{D_{KL}[q(\mathbf{z}|\mathbf{x};\ \phi)||p(\mathbf{z}|\mathbf{x};\ \theta)]}$

事後確率 (p(**z**|**x**; θ))を含むため計算ができない

Evidence Lower Bound

$$
\begin{aligned}
D_{\mathrm{KL}}\left[q(\boldsymbol{z}|\boldsymbol{x})\|p(\boldsymbol{z}|\boldsymbol{x})\right] &= \int q(\boldsymbol{z}|\boldsymbol{x}) \log \frac{q(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z}|\boldsymbol{x})} d\boldsymbol{z} \\
&= \int q(\boldsymbol{z}|\boldsymbol{x}) \left(\log q(\boldsymbol{z}|\boldsymbol{x}) - \underbrace{\boxed{\log \frac{p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{p(\boldsymbol{x})}}}_{\text{Bayes's rule}} \right) d\boldsymbol{z} \\
&= \log p(\boldsymbol{x}) + \int q(\boldsymbol{z}|\boldsymbol{x}) \left(\log \frac{q(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z})} - \log p(\boldsymbol{x}|\boldsymbol{z}) \right) d\boldsymbol{z} \\
&= \log p(\boldsymbol{x}) + D_{KL}\left[q(\boldsymbol{z}|\boldsymbol{x}\|p(\boldsymbol{z}))\right] - \int q(\boldsymbol{z}|\boldsymbol{x}) \log p(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{z}
\end{aligned}
$$

Jensen's inequality

$$
\log p(\boldsymbol{x}) \geq \underline{D_{KL}\left[q(\boldsymbol{z}|\boldsymbol{x}\|p(\boldsymbol{z}))\right] - \int q(\boldsymbol{z}|\boldsymbol{x}) \log p(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{z}}
$$

ELBO

Bayes's rule

$$
p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{p(\boldsymbol{x})}
$$

$$
p(\boldsymbol{x}|\boldsymbol{z}) = \frac{p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{z})}
$$

Kullback-Leibler Divergence

$$\log p(\boldsymbol{x}) \geq \boxed{D_{KL}\left[q(\boldsymbol{z}|\boldsymbol{x}) \| p(\boldsymbol{z})\right)]} - \int q(\boldsymbol{z}|\boldsymbol{x}) \log p(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{z}$$

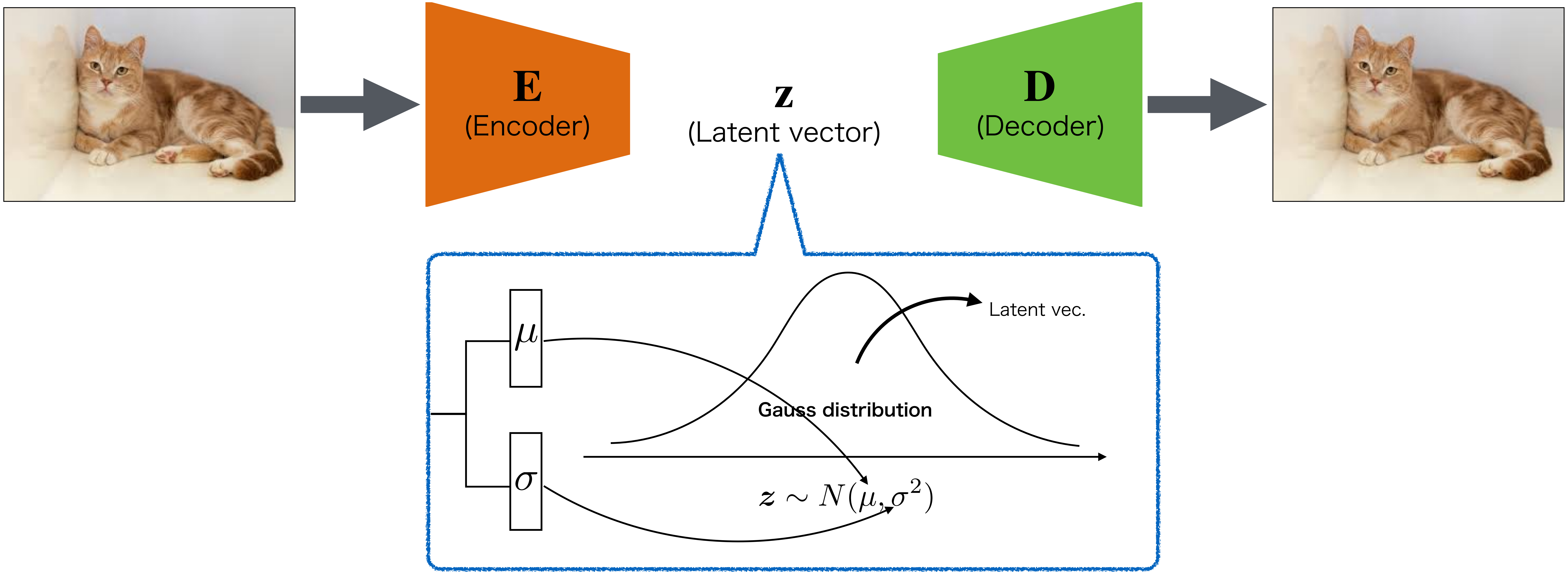
- p(**z**)は多変量正規分布を仮定
- q(**z**|**x**)は多変量正規分布p(**z**)と近似させる

$$\begin{aligned} D_{KL}\left[q(\boldsymbol{z}|\boldsymbol{x}) \| p(\boldsymbol{z})\right] &= D_{KL}\left[N(\boldsymbol{z};\mu,\Sigma) \| N(\boldsymbol{z};0,\boldsymbol{I})\right] \\ &= \frac{1}{2}\left[tr(\Sigma_1^{-1})\Sigma_0 + (\mu_1-\mu_0)^T\Sigma_1^{-1}(\mu_1-\mu_0) - k + \log_e \frac{|\Sigma_1|}{|\Sigma_0|}\right] \\ &= \frac{1}{2}\left(tr(\Sigma_0) + \mu_0^T\mu_0 - k - \log_e |\Sigma_0|\right)\right) \end{aligned}$$

多変量正規分布 (平均: μ, 共分散: Σ)

$$N(\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\right)$$

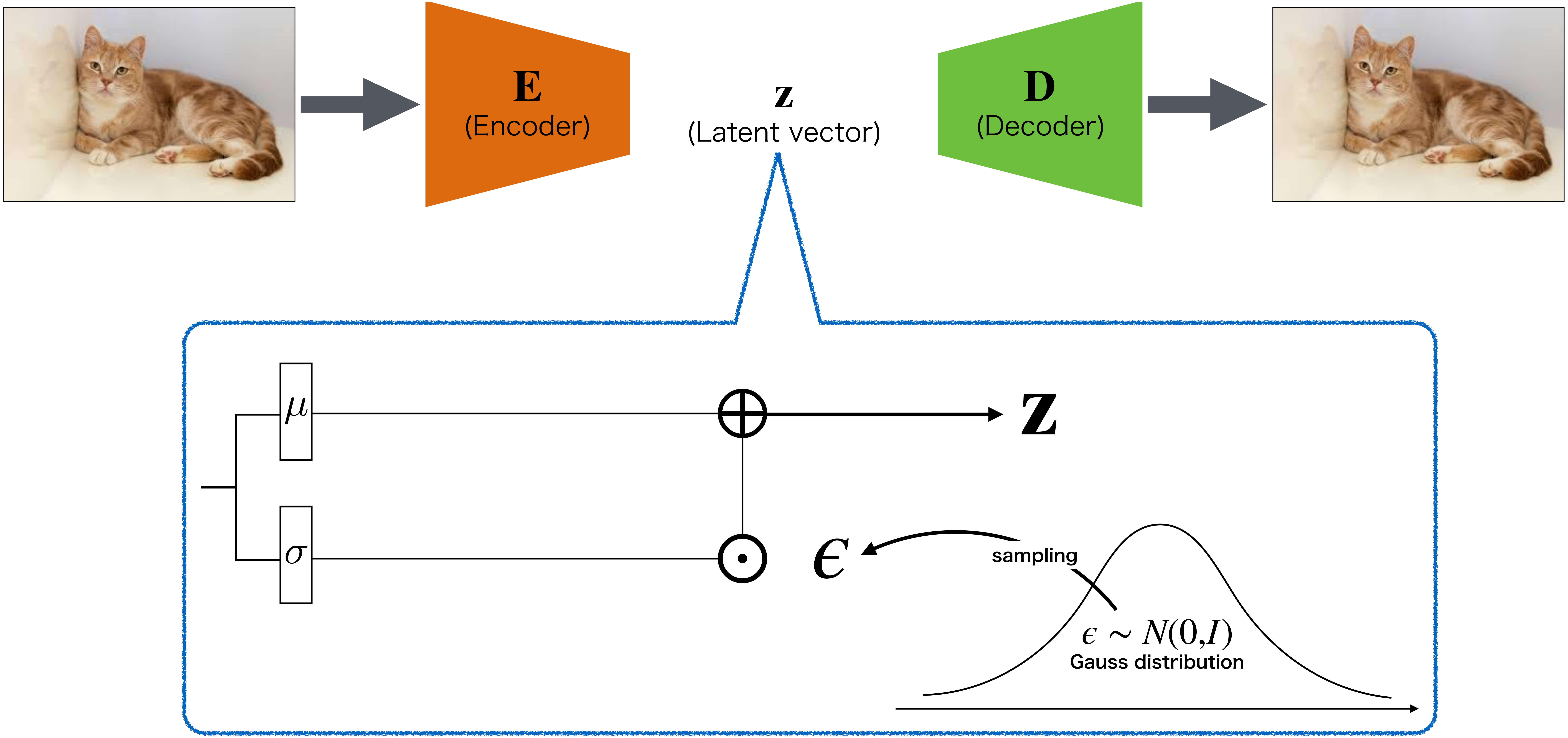
Reparameterization Trick



Reparameterization Trick

This situation cannot use back-propagation

Reparameterization Trick



Reparameterization Trick

VAEを動作させた結果