# 目次

# 目次

# 生成モデル

- 生成モデルの目的
  - データ $\boldsymbol{x}$ に関する分布 $p(\boldsymbol{x})$ を推定する

- データ生成過程
  - データ $\boldsymbol{x}$ は一般に高次元である
  - 但し、実際にデータが分布しているのは、ごく限られた一部の低次元の領域であると考えられる (多様体仮説)
  - データ $\boldsymbol{x}$ 自体は高次元だが、本質的には低次元の情報しか持たないと考えられる

  - データ $\boldsymbol{x}$ を、より低次元なベクトル $\boldsymbol{z}$ を使って、表現することを考える
  - データに関する分布 $p(\boldsymbol{x})$ を、潜在変数 $\boldsymbol{z}$ に関する分布と、うまく組み合わせて記述する
  - 潜在変数からデータが生成されるまでの過程を組み込んで、$p(\boldsymbol{x})$ を記述する

# 目次

1. 変分自己符号化器
   - 生成モデル
   - 変分自己符号化器 (VAE) の概要
   - 変分自己符号化器 (VAE) の理論

# 変分自己符号化器 (VAE) の概要

- 深層学習における生成モデル
  - 主に以下の 2 つの手法が存在する
  - 敵対的生成ネットワーク (Generative Adversarial Networks, GAN)
  - 変分自己符号化器 (Variational Auto Encoders, VAE)
  - ここでは変分自己符号化器 (VAE) について扱う
  - VAE を、異常検知 (不良品の検出など) に使った例がある

# 変分自己符号化器 (VAE) の概要

- VAE におけるグラフィカルモデル
  - 図 1 のような、潜在変数を含んだグラフィカルモデルを考える

  - データ $\boldsymbol{x}$ について、ある一つの潜在変数 $\boldsymbol{z}$ が対応しているとする
  - 各データ $\boldsymbol{x}$ は、分布 $p(\boldsymbol{x})$ から独立にサンプルされるとする
  - 従って、データ $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ は独立同分布標本とする

  - $\theta$ は、潜在変数 $\boldsymbol{z}$ からデータ $\boldsymbol{x}$ を取得する際に使用されるパラメータ
  - $\phi$ は、データ $\boldsymbol{x}$ から潜在変数 $\boldsymbol{z}$ を生成する際に使用されるパラメータ
  - $N$ は、データ数である

# 変分自己符号化器 (VAE) の概要

- データ $\boldsymbol{x}$ の生成過程
  - データ $\boldsymbol{x}$ の生成過程は、次のように考える

  - 分布 $p(\boldsymbol{z}|\theta)$ から、潜在変数 $\boldsymbol{z}_i$ がサンプルされる
  - 分布 $p(\boldsymbol{x}|\boldsymbol{z}_i, \theta)$ から、データ $\boldsymbol{x}_i$ がサンプルされる

  - これより、データ $\boldsymbol{x}$ の分布を次のように表現できる

$$p(\boldsymbol{x}|\theta) = \int p(\boldsymbol{x}|\boldsymbol{z}, \theta)p(\boldsymbol{z}|\theta)d\boldsymbol{z} \tag{1}$$

- 潜在変数 $\boldsymbol{z}$ をデータ $\boldsymbol{x}$ から取得する過程
  - 潜在変数 $\boldsymbol{z}_i$ をデータ $\boldsymbol{x}_i$ から得る過程は、次のように考える

  - 分布 $q(\boldsymbol{z}|\boldsymbol{x}_i, \phi)$ から、潜在変数 $\boldsymbol{z}_i$ がサンプルされる

# 変分自己符号化器 (VAE) の概要

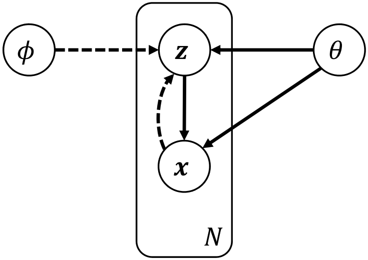


図 1: 変分自己符号化器 (VAE) におけるグラフィカルモデル

# 変分自己符号化器 (VAE) の概要

- 確率分布のニューラルネットワークによる表現
  - 潜在変数を含む確率モデルについて、パラメータの最尤解を求めるために、EM アルゴリズムを導出した
  - EM アルゴリズムでは、潜在変数に関する事後分布 $p(\boldsymbol{z}|\boldsymbol{x}, \theta)$ を計算する必要があった
  - この事後分布 $p(\boldsymbol{z}|\boldsymbol{x}, \theta)$ の計算が困難であるとき、$p(\boldsymbol{z}|\boldsymbol{x})$ を別の分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ で近似し、変分推論によって $q(\boldsymbol{z}|\phi)$ の最適解を求めた

  - VAE は変分推論の変種であり、近似事後分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ と、$p(\boldsymbol{x}|\boldsymbol{z})$ の 2 つをニューラルネットワークで表現する

  - データ $\boldsymbol{x}$ を潜在変数 $\boldsymbol{z}$ に対応付けるニューラルネットワークを、Encoder という
  - 潜在変数 $\boldsymbol{z}$ からデータ $\boldsymbol{x}$ を復元するニューラルネットワークを、Decoder という
  - 分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ は Encoder、分布 $p(\boldsymbol{x}|\boldsymbol{z}, \theta)$ は Encoder に相当する

# 目次

# 変分自己符号化器 (VAE) の理論

- 変分自己符号化器 (VAE) の理論
  - 変分下界 $\mathcal{L}(q)$ は次のようであった

$$
\begin{aligned}
\mathcal{L}(q) &= \int q(\boldsymbol{z}|\boldsymbol{x}) \ln \frac{p(\boldsymbol{x}, \boldsymbol{z})}{q(\boldsymbol{z}|\boldsymbol{x})} d\boldsymbol{z} && (2) \\
&= \int q(\boldsymbol{z}|\boldsymbol{x}) \ln \frac{p(\boldsymbol{x}|\boldsymbol{z}) p(\boldsymbol{z})}{q(\boldsymbol{z}|\boldsymbol{x})} d\boldsymbol{z} && (3) \\
&= \int q(\boldsymbol{z}|\boldsymbol{x}) \ln p(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{z} + \int q(\boldsymbol{z}|\boldsymbol{x}) \ln \frac{p(\boldsymbol{z})}{q(\boldsymbol{z}|\boldsymbol{x})} d\boldsymbol{z} && (4) \\
&= \int q(\boldsymbol{z}|\boldsymbol{x}) \ln p(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{z} - \mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}) || p(\boldsymbol{z})) && (5) \\
&= \mathbb{E}_{\boldsymbol{z} \sim q(\boldsymbol{z}|\boldsymbol{x})} \left[ \ln p(\boldsymbol{x}|\boldsymbol{z}) \right] - \mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}) || p(\boldsymbol{z})) && (6)
\end{aligned}
$$

  - ここでは、単一のデータ $\boldsymbol{x}$ と、それに対応する潜在変数 $\boldsymbol{z}$ を考えている
  - また、パラメータ $\theta, \phi$ は省略している

# 変分自己符号化器 (VAE) の理論

- 第 1 項 $\mathbb{E}_{\boldsymbol{z}\sim q(\boldsymbol{z}|\boldsymbol{x})}[\ln p(\boldsymbol{x}|\boldsymbol{z})]$ を大きく、また第 2 項 $\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z}))$ を小さくすることで、変分下界 $\mathcal{L}(q)$ を大きくできる
- VAE では、変分下界 $\mathcal{L}(q)$ を最大化するパラメータ $\theta, \phi$ を求めるために、ニューラルネットを使用する (変分推論にニューラルネットをねじ込んだもの)
- KL ダイバージェンスの項は、後ほど求めることにする (解析的に求められる)
- 第 1 項は、分布 $q(\boldsymbol{z}|\boldsymbol{x})$ に関する期待値であり、VAE ではサンプリングで近似する

$$\mathbb{E}_{\boldsymbol{z}\sim q(\boldsymbol{z}|\boldsymbol{x})}[\ln p(\boldsymbol{x}|\boldsymbol{z})] \simeq \frac{1}{L}\sum_{l=1}^{L} \ln p(\boldsymbol{x}_i|\boldsymbol{z}_{i,l}) \tag{7}$$

# 変分自己符号化器 (VAE) の理論

- これより、変分下界 $\mathcal{L}(q)$ は以下のように書ける

$$\mathcal{L}(q) \simeq -\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z})) + \frac{1}{L}\sum_{i=1}^{L} \ln p(\boldsymbol{x}_i|\boldsymbol{z}_{i,l}) \tag{8}$$

- パラメータ $\theta, \phi$ を含めれば、次のように書ける

$$\mathcal{L}(q) \simeq -\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}, \phi)||p(\boldsymbol{z}|\theta)) + \frac{1}{L}\sum_{i=1}^{L} \ln p(\boldsymbol{x}_i|\boldsymbol{z}_{i,l}, \theta) \tag{9}$$

# 変分自己符号化器 (VAE) の理論

- Encoder のニューラルネットの入出力
  - Encoder は、分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ を表現するニューラルネット
  - 入力 $\boldsymbol{x}$ に対応する潜在変数 $\boldsymbol{z}$ を得る

  - Encoder の入力は、明らかにデータ $\boldsymbol{x}$ である

  - 変分下界 $\mathcal{L}(q)$ において、項 $\mathbb{E}_{\boldsymbol{z}}\left[\ln p(\boldsymbol{x}|\boldsymbol{z})\right]$ は近似する必要があった
  - $\boldsymbol{z}$ を、分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ から $L$ 回サンプリングした

  - ニューラルネットで、$\boldsymbol{x}$ から $\boldsymbol{z}$ を直接サンプリングするのは困難
  - そこで、Encoder では、サンプルされたデータは出力しないことにする
  - その代わりに、サンプルする分布のパラメータを出力する

  - 例えば、サンプルする分布がガウス分布であれば、平均と分散の 2 つのパラメータを出力する

# 変分自己符号化器 (VAE) の理論

- 後述のように、分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ はガウス分布になるので、Encoder は平均ベクトル $\boldsymbol{\mu}$ と共分散行列 $\boldsymbol{\Sigma}$ を出力する
- 共分散行列は、実際には対角行列であるため、実際には行列ではなく、行列の対角成分を要素にもつベクトルを出力する

# 変分自己符号化器 (VAE) の理論

- Encoder の損失関数
  - VAE では、事前分布として、平均ベクトル $\mathbf{0}$、共分散行列 $\boldsymbol{I}$ のガウス分布 $\mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})$ を仮定する
  - データ $\boldsymbol{x}$ は高次元だが、実際にはそのうちの低次元な領域にまとまって存在する (多様体仮説)
  - 従って、データ $\boldsymbol{x}$ の構造を、より低次元な潜在変数 $\boldsymbol{z}$ の空間 $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})$ に押し込めることができる
  - Encoder の損失関数は、KL ダイバージェンス $\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}, \phi)||p(\boldsymbol{z}|\theta))$ で定義できる
  - この KL ダイバージェンスを最小化することは、Encoder の分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ を、$p(\boldsymbol{z}|\theta) \sim \mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})$ に近づける制約に相当する
  - $p(\boldsymbol{z}|\theta)$ がガウス分布であれば、事後分布 $p(\boldsymbol{z}|\boldsymbol{x}, \theta)$ もガウス分布であり、従って $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ もガウス分布となる

# 変分自己符号化器 (VAE) の理論

- よって $\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x},\phi)||p(\boldsymbol{z}|\theta))$ は、2 つのガウス分布間の KL ダイバージェンスである

- 一般に、2 つのガウス分布 $p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$、$q(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ 間の KL ダイバージェンスは、解析的に計算できる
- KL ダイバージェンス $\mathrm{KL}(p(\boldsymbol{z})||q(\boldsymbol{z}))$ を順番に求めてみよう

$$
\begin{aligned}
& \mathrm{KL}(p(\boldsymbol{z})||q(\boldsymbol{z})) \\
= & \int p(\boldsymbol{z}) \ln \frac{p(\boldsymbol{z})}{q(\boldsymbol{z})} d\boldsymbol{z} && (10) \\
= & \int p(\boldsymbol{z}) \left( \ln p(\boldsymbol{z}) - \ln q(\boldsymbol{z}) \right) d\boldsymbol{z} && (11) \\
= & \int p(\boldsymbol{z}) \left( \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) - \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \right) d\boldsymbol{z} && (12) \\
= & \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})} \left[ \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) - \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \right] && (13)
\end{aligned}
$$

- データを $D$ 次元、潜在変数を $K$ 次元とする

- ここで

$$
\begin{aligned}
& \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\
= \quad & \ln \left( \frac{1}{(2\pi)^{\frac{K}{2}}} \frac{1}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \exp\left( -\frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_0) \right) \right) \\
= \quad & -\frac{K}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_0| - \frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_0) \qquad (14)
\end{aligned}
$$

$$
\begin{aligned}
& \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\
= \quad & -\frac{K}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_1| - \frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_1) \qquad (15)
\end{aligned}
$$

であるので

$$
\begin{aligned}
& \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) - \ln \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\
= \quad & -\frac{K}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_0| - \frac{1}{2} (\boldsymbol{z} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_0) -
\end{aligned}
$$

$$
\begin{aligned}
& \left(-\frac{K}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_1| - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)\right) \\
= \quad & \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_0) + \\
& \quad \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)
\end{aligned} \tag{16}
$$

- よって

$$
\begin{aligned}
& \mathrm{KL}(p(\boldsymbol{z})||q(\boldsymbol{z})) \\
= \quad & \mathbb{E}_{\boldsymbol{z}\sim p(\boldsymbol{z})}\left[\ln\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0) - \ln\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)\right] \\
= \quad & \mathbb{E}_{p(\boldsymbol{z})}\left[\frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_0) + \right. \\
& \quad \left. \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)\right] \qquad (17) \\
= \quad & \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}\mathbb{E}_{p(\boldsymbol{z})}\left[(\boldsymbol{z}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_0)\right] +
\end{aligned}
$$

# 変分自己符号化器 (VAE) の理論

$$\frac{1}{2}\mathbb{E}_{p(\boldsymbol{z})}\left[\left(\boldsymbol{z}-\boldsymbol{\mu}_1\right)^T\boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{z}-\boldsymbol{\mu}_1\right)\right] \tag{18}$$

- ここで、期待値についての式を導出しておく
- $\mathbb{E}\left[\boldsymbol{z}\right]=\boldsymbol{\mu}$、$\mathbb{E}\left[\left(\boldsymbol{z}-\boldsymbol{\mu}\right)\left(\boldsymbol{z}-\boldsymbol{\mu}\right)^T\right]=\boldsymbol{\Sigma}$ とする
- $\boldsymbol{z}$ の $i$ 成分を $z_i$、$\boldsymbol{\mu}$ の $i$ 成分を $\mu_i$、$\boldsymbol{\Sigma}$ の $i,j$ 成分を $\Sigma_{ij}$ とする
- このとき

$$\begin{aligned}
\Sigma_{ij} &= \mathbb{E}\left[\left(z_i-\mu_i\right)\left(z_j-\mu_j\right)\right] \\
&= \mathbb{E}\left[z_iz_j-z_i\mu_j-z_j\mu_i+\mu_i\mu_j\right] \\
&= \mathbb{E}\left[z_iz_j\right]-\mu_j\mathbb{E}\left[z_i\right]-\mu_i\mathbb{E}\left[z_j\right]+\mu_i\mu_j \\
&= \mathbb{E}\left[z_iz_j\right]-\mu_j\mu_i-\mu_i\mu_j+\mu_i\mu_j \\
&= \mathbb{E}\left[z_iz_j\right]+\mu_i\mu_j
\end{aligned} \tag{19}$$

であるから

$$\mathbb{E}\left[z_iz_j\right]=\Sigma_{ij}-\mu_i\mu_j \tag{20}$$

# 変分自己符号化器 (VAE) の理論

- そして、行列 $\boldsymbol{A}$ の $i, j$ 成分を $A_{ij}$ とすれば、以下を得る

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z}\right] &= \mathbb{E}\left[\sum_i\sum_j z_i A_{ij} z_j\right] \\
&= \sum_i\sum_j A_{ij}\mathbb{E}\left[z_i z_j\right] \\
&= \sum_i\sum_j A_{ij}\left(\Sigma_{ij} + \mu_i\mu_j\right) \\
&= \sum_i\sum_j A_{ij}\Sigma_{ij} + \sum_i\sum_j A_{ij}\mu_i\mu_j \\
&= \sum_i\sum_j A_{ij}\Sigma_{\color{red}{ji}} + \sum_i\sum_j \mu_i A_{ij}\mu_j \\
&= \sum_i\left(\boldsymbol{A}\boldsymbol{\Sigma}\right)_{ii} + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu} \\
&= \mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{\Sigma}\right) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu} \qquad (21)
\end{aligned}
$$

# 変分自己符号化器 (VAE) の理論

- 上式の変形では、共分散行列 $\mathbf{\Sigma}$ が対称行列ゆえ、$\Sigma_{ij} = \Sigma_{ji}$ が成立することを用いた
- また、ベクトル $\boldsymbol{a}$ の $i$ 成分を $a_i$ とすれば、以下を得る

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{a}^T \boldsymbol{z}\right] &= \mathbb{E}\left[\boldsymbol{z}^T \boldsymbol{a}\right] \\
&= \mathbb{E}\left[\sum_i z_i a_i\right] \\
&= \sum_i a_i \mathbb{E}\left[z_i\right] \\
&= \sum_i a_i \mu_i \\
&= \boldsymbol{a}^T \boldsymbol{\mu} = \boldsymbol{\mu}^T \boldsymbol{a} \qquad (22)
\end{aligned}
$$

# 変分自己符号化器 (VAE) の理論

- これより、$\boldsymbol{a}, \boldsymbol{B}$ をそれぞれ適当なベクトル、行列とすれば、以下を得る

$$
\begin{aligned}
& \mathbb{E}\left[(\boldsymbol{z}-\boldsymbol{a})^{T} \boldsymbol{B}(\boldsymbol{z}-\boldsymbol{a})\right] \\
= & \mathbb{E}\left[\boldsymbol{z}^{T} \boldsymbol{B} \boldsymbol{z}-\boldsymbol{z}^{T} \boldsymbol{B} \boldsymbol{a}-\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{z}+\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{a}\right] \\
= & \mathbb{E}\left[\boldsymbol{z}^{T} \boldsymbol{B} \boldsymbol{z}\right]-\mathbb{E}\left[\boldsymbol{z}^{T} \boldsymbol{B} \boldsymbol{a}\right]-\mathbb{E}\left[\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{z}\right]+\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{a} \\
= & \left(\operatorname{Tr}(\boldsymbol{B} \boldsymbol{\Sigma})+\boldsymbol{\mu}^{T} \boldsymbol{B} \boldsymbol{\mu}\right)-\boldsymbol{\mu}^{T} \boldsymbol{B} \boldsymbol{a}-\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{\mu}+\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{a} \\
= & \operatorname{Tr}(\boldsymbol{B} \boldsymbol{\Sigma})+\boldsymbol{\mu}^{T} \boldsymbol{B} \boldsymbol{\mu}-2 \boldsymbol{\mu}^{T} \boldsymbol{B} \boldsymbol{a}+\boldsymbol{a}^{T} \boldsymbol{B} \boldsymbol{a}
\end{aligned} \tag{23}
$$

特に、$\boldsymbol{a}=\boldsymbol{\mu}, \boldsymbol{B}=\boldsymbol{\Sigma}^{-1}$ とすれば

$$
\begin{aligned}
& \mathbb{E}\left[(\boldsymbol{z}-\boldsymbol{\mu})^{T} \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right] \\
= & \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}\right)+\boldsymbol{\mu}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}-2 \boldsymbol{\mu}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}+\boldsymbol{\mu}^{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
= & \operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}\right) \\
= & \operatorname{Tr}(\boldsymbol{I})=K
\end{aligned} \tag{24}
$$

# 変分自己符号化器 (VAE) の理論

- これを用いれば、KL ダイバージェンスは次のようになる

$$
\begin{aligned}
&\mathrm{KL}(p(\boldsymbol{z})||q(\boldsymbol{z})) \\
= \quad & \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}\mathbb{E}_{p(\boldsymbol{z})}\left[\left(\boldsymbol{z}-\boldsymbol{\mu}_0\right)^T \boldsymbol{\Sigma}_0^{-1}\left(\boldsymbol{z}-\boldsymbol{\mu}_0\right)\right] + \\
& \quad \frac{1}{2}\mathbb{E}_{p(\boldsymbol{z})}\left[\left(\boldsymbol{z}-\boldsymbol{\mu}_1\right)^T \boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{z}-\boldsymbol{\mu}_1\right)\right] \\
= \quad & \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}K + \frac{1}{2}\left(\mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right) + \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_0 - \right. \\
& \quad \left. 2\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1\right) \\
= \quad & \frac{1}{2}\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - \frac{1}{2}K + \frac{1}{2}\,\mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right) + \\
& \quad \frac{1}{2}\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right)^T \boldsymbol{\Sigma}_1^{-1}\left(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1\right) \\
= \quad & \frac{1}{2}\left(\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} - K + \mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right) + \right.
\end{aligned}
$$

# 変分自己符号化器 (VAE) の理論

$$\left.(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)\right) \tag{25}$$

- これで、2 つのガウス分布 $p(\boldsymbol{z})=\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$、$q(\boldsymbol{z})=\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$ 間の KL ダイバージェンスが、次のようになることが分かった

$$\begin{aligned}&\mathrm{KL}(p(\boldsymbol{z})||q(\boldsymbol{z}))\\ =\;&\frac{1}{2}\left(\ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|}-K+\mathrm{Tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\right)+\right.\\ &\left.\qquad(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)\right)\end{aligned} \tag{26}$$

- ここでは、$p(\boldsymbol{z})=\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ は、Encoder のニューラルネットが表現する分布 $q(\boldsymbol{z}|\boldsymbol{x},\phi)$ である
- そして、分布 $q(\boldsymbol{z}|\boldsymbol{x},\phi)$ はガウス分布であったので、ここでは平均 $\boldsymbol{\mu}_0$ と共分散行列 $\boldsymbol{\Sigma}_0$ を使って、$q(\boldsymbol{z}|\boldsymbol{x},\phi)=\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0,\boldsymbol{\Sigma}_0)$ と表すことにする

# 変分自己符号化器 (VAE) の理論

- また $q(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ は、潜在変数に関する事後分布 $p(\boldsymbol{z}|\theta) = \mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})$ である
- 結局、KL ダイバージェンス $\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}, \phi)||p(\boldsymbol{z}|\theta))$ は、先程の式に $\boldsymbol{\mu}_1 = \mathbf{0}$ と $\boldsymbol{\Sigma}_1 = \boldsymbol{I}$ を代入すれば得られる

$$
\begin{aligned}
& \mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x}, \phi)||p(\boldsymbol{z}|\theta)) \\
= \; & \mathrm{KL}(\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)||\mathcal{N}(\boldsymbol{z}|\mathbf{0}, \boldsymbol{I})) \\
= \; & \frac{1}{2}\left(-\ln|\boldsymbol{\Sigma}_0| - K + \mathrm{Tr}\left(\boldsymbol{\Sigma}_0\right) + \boldsymbol{\mu}_0^T \boldsymbol{\mu}_0\right)
\end{aligned} \tag{27}
$$

- この KL ダイバージェンスが、Encoder の損失関数として定義される
- Encoder のニューラルネットは、入力としてデータ $\boldsymbol{x}$ を取り、平均 $\boldsymbol{\mu}_0$ と共分散行列 $\boldsymbol{\Sigma}_0$ を出力する
- 従って、Encoder の出力と、潜在変数の次元 $K$ を上式に代入すれば、損失関数を容易に計算できる

# 変分自己符号化器 (VAE) の理論

- $\boldsymbol{\Sigma}_0$ は対称行列であるため、実際に出力されるのは、$\boldsymbol{\Sigma}_0$ の対角成分を並べたベクトルである
- 一般的な VAE の Encoder は次の図 2 のように表せる

# 変分自己符号化器 (VAE) の理論

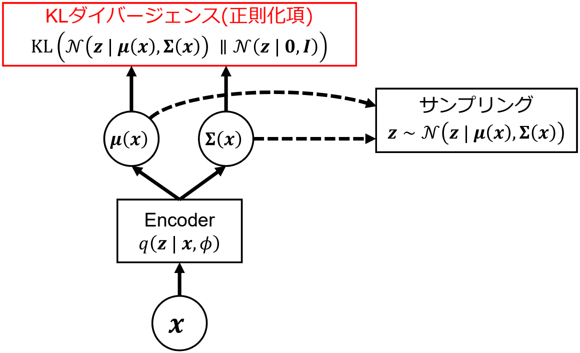


図 2: VAE の Encoder の概要

# 変分自己符号化器 (VAE) の理論

- Encoder のニューラルネットの構造
  - VAE が最初に提案された論文では、隠れ層は 1 層となっている
  - 最終層は、$\boldsymbol{\mu}_0$ と $\sqrt{\boldsymbol{\Sigma}_0}$ を出力する 2 つのユニットから成る
  - $\sqrt{\boldsymbol{\Sigma}_0}$ とは、行列 $\boldsymbol{\Sigma}_0$ の各要素の平方根を取った行列である

  - 隠れ層の重みを $\boldsymbol{W}_h$、バイアスを $\boldsymbol{b}_h$、活性化関数を $f(\cdot)$、層の出力を $\boldsymbol{h}$ とする
  - 隠れ層で行う処理は、次の式で表される

$$\boldsymbol{h} = f(\boldsymbol{W}_h \boldsymbol{x} + \boldsymbol{b}_h) \tag{28}$$

  - $\boldsymbol{\mu}_0$ は、重み $\boldsymbol{W}_m$ とバイアス $\boldsymbol{b}_m$ を使って、以下のように計算される

$$\boldsymbol{\mu}_0 = \boldsymbol{W}_m \boldsymbol{h} + \boldsymbol{b}_m \tag{29}$$

  - $\sqrt{\boldsymbol{\Sigma}_0}$ は、重み $\boldsymbol{W}_s$ とバイアス $\boldsymbol{b}_s$ から、以下のように計算される

$$\sqrt{\boldsymbol{\Sigma}_0} = \boldsymbol{W}_s \boldsymbol{h} + \boldsymbol{b}_s \tag{30}$$

# 変分自己符号化器 (VAE) の理論

- Decoder のニューラルネットの入出力
  - Decoder は、分布 $p(\boldsymbol{x}|\boldsymbol{z},\theta)$ を表現するニューラルネット
  - 潜在変数 $\boldsymbol{z}$ から元のデータ $\boldsymbol{x}$ を復元する

  - Decoder の入力は、Encoder によってサンプリングされた $\boldsymbol{z}$ となる

  - もう少し正確に表現すると、Encoder から出力されるのは、分布のパラメータ $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ である
  - そして、そのパラメータを使って分布 $\mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ を構成し、分布から $\boldsymbol{z}$ をサンプリングする

  - Decoder の出力は、復元されたデータ $\boldsymbol{y}$ である

# 変分自己符号化器 (VAE) の理論

- Decoder の損失関数
  - 画像データでは通常、各ピクセルの値が $0$ から $1$ までになるようにスケーリングされている
  - このとき、$p(\boldsymbol{x}|\boldsymbol{z})$ はベルヌーイ分布と仮定していることになる

  - 出力層のユニット $j$ は、$y_j = p(x_j = 1|\boldsymbol{z})$ を出力しているとみなせる
  - $y_j$ は、再構成 $\boldsymbol{y}$ の $j$ 番目の要素であり、元のデータ $\boldsymbol{x}$ の $j$ 番目の要素 $x_j$ と対応する

  - VAE が最初に提案された論文では、隠れ層は 1 層のみである
  - 再構成 $\boldsymbol{y}$ は、次のように計算される

$$\boldsymbol{y} = f_\sigma\left(\boldsymbol{W}_o \tanh\left(\boldsymbol{W}_h \boldsymbol{z} + \boldsymbol{b}_h\right) + \boldsymbol{b}_o\right) \tag{31}$$

  - $f_\sigma(\cdot)$ は、行列の各要素にシグモイド関数 $\sigma(\cdot)$ を適用する活性化関数
  - $\boldsymbol{W}_h, \boldsymbol{b}_h$ は隠れ層の重みとバイアス、$\boldsymbol{W}_o, \boldsymbol{b}_o$ は出力層の重みとバイアス

# 変分自己符号化器 (VAE) の理論

- このとき $\ln p(\boldsymbol{x}|\boldsymbol{z})$ は次のように記述できる

$$
\begin{aligned}
\ln p(\boldsymbol{x}|\boldsymbol{z}) &= \ln \prod_{j=1}^{D} \left(p(x_j = 1|\boldsymbol{z})\right)^{x_j} \left(p(x_j = 0|\boldsymbol{z})\right)^{1-x_j} \\
&= \ln \prod_{j} \left(p(x_j = 1|\boldsymbol{z})\right)^{x_j} \left(1 - p(x_j = 1|\boldsymbol{z})\right)^{1-x_j} \\
&= \ln \prod_{j} y_j^{x_j} \left(1 - y_j\right)^{1-x_j} \\
&= \sum_{j} \left(x_j \ln y_j + (1 - x_j) \ln (1 - y_j)\right) \qquad (32)
\end{aligned}
$$

- $\boldsymbol{z}$ は、分布 $q(\boldsymbol{z}|\boldsymbol{x}, \phi)$ からサンプリングされている
- 従って、上記を最大化することは、$\mathbb{E}_{q(\boldsymbol{z})}\left[\ln p(\boldsymbol{x}|\boldsymbol{z})\right]$ を最大化することに等しい

# 変分自己符号化器 (VAE) の理論

- 上記は、$x_j$ と $y_j$ のいずれもベルヌーイ分布に従う (二値変数) ときの、負の交差エントロピーとなっていることが分かる
- 従って、$\mathbb{E}_{q(\boldsymbol{z})}[\ln p(\boldsymbol{x}|\boldsymbol{z})]$ を最大化することは、交差エントロピーを最小化することに相当

- Decoder の損失関数は、以下の交差エントロピー誤差として定義できる

$$E = -\sum_{j} (x_j \ln y_j + (1 - x_j) \ln (1 - y_j)) \tag{33}$$

- 元データ $x_j$ と、その再構成 $y_j$ との差が大きければ大きいほど、上記の誤差は増大する
- これより、上記の誤差は再構成誤差 (Reconstruction Error) とよばれる

- これより、VAE の Encoder と Decoder は次の図 3 のように表せる

# 変分自己符号化器 (VAE) の理論

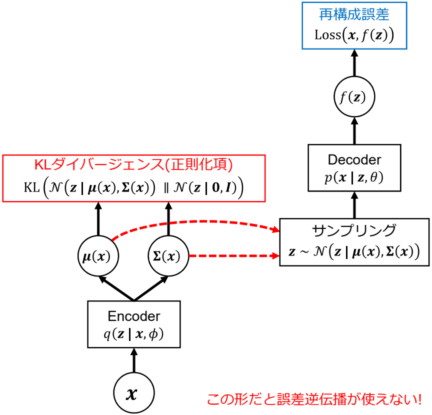


図 3: VAE の Encoder と Decoder の概要

# 変分自己符号化器 (VAE) の理論

- Reparameterization Trick
  - VAE が先程の図 3 のようであるとき、大きな問題が生じる
  - サンプリングを行うと、計算グラフが途中で途切れるため、誤差逆伝播法を実行できない

  - そこで、次の図 4 のように構成する
  - $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ として、$\boldsymbol{z}$ を分布から直接サンプリングするのではない
  - $\boldsymbol{z}$ を、決定論的な関数 $g(\boldsymbol{\epsilon}, \boldsymbol{x}|\phi)$ から決定する
  - 但し、$\boldsymbol{\epsilon}$ は、分布 $p(\boldsymbol{\epsilon})$ からサンプリングされる

  - ニューラルネットの最適化には無関係な項 $\boldsymbol{\epsilon}$ と、Encoder のパラメータ $\phi$ で $\boldsymbol{z}$ を表現することで、誤差逆伝播法を実行可能にする
  - このテクニックを、Reparameterization Trick という

  - $\boldsymbol{\epsilon} \sim (\boldsymbol{\epsilon}|\boldsymbol{0}, \boldsymbol{I})$ とすれば、$\boldsymbol{z}$ は次のように計算できる

$$\boldsymbol{z} = g(\boldsymbol{\epsilon}, \boldsymbol{x}|\phi) = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon} \tag{34}$$

# 変分自己符号化器 (VAE) の理論

- 共分散行列 $\mathbf{\Sigma}$ が対角行列であれば、上記の $\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon}$ は、単なる要素ごとの積 (対角行列の各要素を並べたベクトルと、$\boldsymbol{\epsilon}$ の要素ごとの積) として書ける
- $\boldsymbol{z}$ の式は以下のように導出できる
- 確率変数 $\boldsymbol{z}, \boldsymbol{\epsilon}$ 間の関係が、次のようになっているとする

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{U}\mathbf{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon} \tag{35}$$

- 但し、正定値対称行列 $\mathbf{\Sigma}$ が、固有値分解によって $\mathbf{\Sigma} = \boldsymbol{U}\mathbf{\Lambda}\boldsymbol{U}^T = \boldsymbol{U}\mathbf{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{U}\mathbf{\Lambda}\right)^T$ と表せるとする
- $\boldsymbol{U}$ は固有ベクトルを並べた行列、$\mathbf{\Lambda}$ は対角成分に固有値をもつ対角行列とする

# 変分自己符号化器 (VAE) の理論

- 確率分布 $p(\boldsymbol{z})$ と $p(\boldsymbol{\epsilon})$ との関係は、ヤコビ行列 $\boldsymbol{J} = \partial\boldsymbol{\epsilon}/\partial\boldsymbol{z}$ により次のように記述できる

$$p(\boldsymbol{z}) = p(\boldsymbol{\epsilon})|\det(\boldsymbol{J})| = p(\boldsymbol{\epsilon})\left|\det\left(\frac{\partial\boldsymbol{z}}{\partial\boldsymbol{\epsilon}}\right)\right| \tag{36}$$

- ヤコビ行列 $\boldsymbol{J}$ を計算すると次のようになる

$$\boldsymbol{J} = \frac{\partial\boldsymbol{\epsilon}}{\partial\boldsymbol{z}} = \frac{\partial}{\partial\boldsymbol{z}}\left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right) = \left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\right)^{T} \tag{37}$$

# 変分自己符号化器 (VAE) の理論

- ヤコビ行列 $\boldsymbol{J}$ の行列式は次のようになる

$$\begin{aligned}\det(\boldsymbol{J}) &= \left|\left(\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\right)^{T}\right| \\ &= \left|\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}\right| \qquad (\because |\boldsymbol{A}^{T}| = |\boldsymbol{A}|) \\ &= \frac{1}{\left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right|} \qquad \left(\because |\boldsymbol{A}^{-1}| = \frac{1}{|\boldsymbol{A}|}\right)\end{aligned} \tag{38}$$

- $\boldsymbol{\Sigma}$ の行列式は次のように表せる

$$|\boldsymbol{\Sigma}| = \left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{T}\right| = \left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right|\left|\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{T}\right| = \left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right|^{2} \tag{39}$$

- 従って $\left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right| = |\boldsymbol{\Sigma}|^{\frac{1}{2}}$ である

# 変分自己符号化器 (VAE) の理論

- $\boldsymbol{\Sigma}$ は正定値である ($|\boldsymbol{\Sigma}| > 0$) から、$|\boldsymbol{\Sigma}|^{\frac{1}{2}} = \left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right| > 0$ が成立し、$\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}$ も正定値行列となる
- これより、逆行列 $\left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{-1}$ が存在するので、ヤコビ行列 $\boldsymbol{J}$ は計算できることが確認される

- ヤコビ行列 $\boldsymbol{J}$ の行列式の絶対値は、次のようになる

$$|\det(\boldsymbol{J})| = \left|\frac{1}{\left|\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\right|}\right| = \left|\frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\right| = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \tag{40}$$

- $p(\boldsymbol{\epsilon})$ がガウス分布 $\mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \boldsymbol{I})$ であるとすると、$p(\boldsymbol{z})$ は次のようになる

$$\begin{aligned} p(\boldsymbol{z}) &= p(\boldsymbol{\epsilon}) \left|\det\left(\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{\epsilon}}\right)\right| \\ &= \frac{1}{(2\pi)^{\frac{K}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\right) \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \end{aligned} \tag{41}$$

# 変分自己符号化器 (VAE) の理論

- ここで、指数関数の中身は次のように書ける

$$
\begin{aligned}
& (\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu}) \\
= & \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon}\right)^T \left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T\right)^{-1} \left(\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon}\right) \\
= & \boldsymbol{\epsilon}^T \left(\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^T \boldsymbol{U}^T \left(\boldsymbol{U}^T\right)^{-1} \boldsymbol{\Lambda}^{-1}\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon} \\
= & \boldsymbol{\epsilon}^T \left(\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^T \boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon} \\
= & \boldsymbol{\epsilon}^T \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon} \\
= & \boldsymbol{\epsilon}^T \boldsymbol{\epsilon}
\end{aligned}
\tag{42}
$$

# 変分自己符号化器 (VAE) の理論

- これより、$p(\boldsymbol{z})$ は平均 $\boldsymbol{\mu}$、共分散行列 $\boldsymbol{\Sigma}$ のガウス分布である

$$
\begin{aligned}
p(\boldsymbol{z}) &= \frac{1}{(2\pi)^{\frac{K}{2}}} \exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\right)\frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \\
&= \frac{1}{(2\pi)^{\frac{K}{2}}}\frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\right) \qquad (43) \\
&= \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \qquad (44)
\end{aligned}
$$

- $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ について、共分散行列 $\boldsymbol{\Sigma}$ が既に対角行列であれば、$\boldsymbol{U}=\boldsymbol{I}$、$\boldsymbol{\Lambda}=\boldsymbol{\Sigma}$ であるので、結局以下が言える

$$
\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\epsilon} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\epsilon} \qquad (45)
$$

# 変分自己符号化器 (VAE) の理論

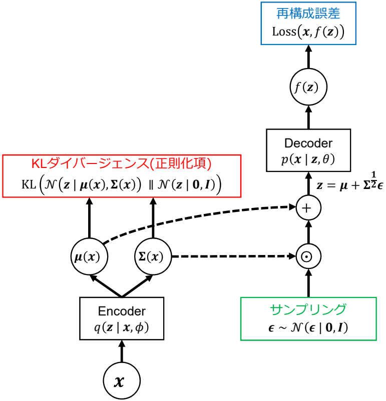


図 4: VAE の構造