VAE□Variational Autoencoder □□□□□□□

- **□AE□□□□□□□□□□□VAE□decoder□□□□□□□□□□□□□□□**

□□□□(AE)□VAE□□□

□□□□(AE)□□□□

- □□□□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□□□
  - □□□□ → □□□□□□ → □□□□□□
- □□□□□□□□□□□□□
- □□□□□□□□□□

VAE□□□

- VAE□□□□□□□□□□□□□□□
- □□□□□□□□□□□□□(□□□□□□□)□□□□□□□(μ)□□□□□($\sigma ^2$)□□
- □□□□□□□□□□□□□□□□□□□□□□□□□□
- □□□□□□□□□VAE□□□

□□□□

- $P(x)$□□□□□□□□□□□□

- z□x□□□□□□□□□□□□□□□z□□□$P(x)$□□□□x□□□□□

  - □□□□□z□□□□□□□□□□z□□□□□□□□□□□□□□□□□□□□□

- VAE□□□□□□□□□□□□□□□□□□**P(x)**

  - $P(x) = \int P(z)P(x|z)$
  - z□□□□□□□□(□□□□□□□□□□$N(0,I))$
  - □□□□z□□□□□□□□□□□□x□□□□□□□□□
  - □□□$P(x)$□□□□□□□□□□□□□□□(□□)

image-20250430233601234

- **Decoder**□□□□□□z□□□□□□□□□□□□(NN)□□□□□□$P(x|z)$□□□□□□□$\mu(z)$□□□$\sigma(z)$□□□□□ □□z□x□□□□□**$P(x|z)$**
- **Encoder**□□□□□□x□□□□□□□□□□□□□□□(NN')□□□□□□□□$q(z|x)$□□□□□□□$\mu'(x)$□□□$\sigma'(x)$□ □□□□□□□□□□□□□□□□□□**$P(z|x)$**image-20250501000147212

□□□□

□□□□□□□[1. □□□□□□□Variational Autoencoder□ — □□□□□□ □□□ □□]

> ***X***□□□□□□□***Observed variable***□□***Z*** □□□□□□□□***Unobserved variable***□
>
> ***X***□□□□□□***Z***□□□□□□□□□□□
>
> - *□X□Z□□□□□X□Z□□□□□P(Z|X)*
> - *□Z□X□□□□□Z□X□□□□□P(X|Z)*
>
> *image-20250502115738249*

□□□□□□□□□□□□□□□□□□ **$P(X,Z)$**

□□□$\mathcal D$□□□□□□□□□□□□□□***X***□***Z***□□□□□□□□□□$\mathcal {D} ={(z^{(1)},x^{(1)}),\cdots, (z^{(N)},x^{(N)})}$,□□□□□□□□□ $$ \mathcal {L}(\theta;\mathcal {D}) = \sum_{i=1}^N lnp_{\theta}(z^{(i)},x^{(i)}) \tag{1.1} $$ □□□□□□□***Z***□□□□□□□□□□□□□□□$\mathcal {D} = {x^{(1)},\cdots,x^{(N)}}$□□□□□□□□□ $$ \mathcal {L}(\theta;\mathcal {D}) = \sum_{i=1}^N \ln{p_{\theta}(x^{(i)})} = \sum_{i=1}^N \ln{\int_{z}{p_{\theta}(x^{(i)},z)dz}} \tag{1.2} $$ □□□□□□□□□□□□

□□□□

□□$Jenson$□□□ $$ \text{□□□□□ f(x)□□□□□□} \ f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)] \ \text{□□□□□□□□□□□□□□} $$ □□□□□***Z***□□□□□□□$q_{\phi}(z|x)$

□□□□□(1.2)□□□□□ $$ \begin{align} \mathcal {L}(\theta;x) &= \ln p_{\theta}(x) \ &=\ln \int_z{p_{\theta}(x,z)} \ &=\ln \int_z{q_{\phi}(z|x)\frac{p_{\theta}(x,z)}{q_{\phi}(z|x)}} \quad \text{□□□□}q_{\phi}(z|x)\text{□□□□□□}\ &= \ln \mathbb{E}{q{\phi}(z|x)}[\frac{p_{\theta}(x,z)}{q_{\phi}(z|x)}] \ &\geq \mathbb{E}{q{\phi}(z|x)} \ln{[\frac{p_{\theta}(x,z)}{q_{\phi}(z|x)}]} \quad \text{□□Jensen□□□} \ &=\int_z{q_{\phi}(z|x)}\ln{[\frac{p_{\theta}(x,z)}{q_{\phi}(z|x)}]} \ &=[\int_z{q_{\phi}(z|x)\ln{p_{\theta}(x,z)}} - \int_z{q_{\phi}(z|x)}\ln{q_{\phi}(z|x)}] \ &\triangleq \mathcal{L}(q,\theta) \end{align} $$ □□□□□□ELBO□$ \mathcal{L}(q,\theta)$ $$ \begin{align} \mathcal{L}(q,\theta) &=[\int_z{q_{\phi}(z|x)\ln{p_{\theta}(x,z)}} - \int_z{q_{\phi}(z|x)}\ln{q_{\phi}(z|x)}] \ &= \mathbb{E}{z \sim q{\phi}}[\ln{p_{\theta}(x,z)}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z|x)}] \end{align} $$ □$p(x,z) = p(z)p(x|z) = p(x)p(z|x)$□□□□□□□□□□□□□□□**ELBO**

□□□□

□□z□□□$p(z|x)$□□□□ $$ \begin{align} \mathcal{L}(q,\theta) &= \mathbb{E}{z \sim q{\phi}}[\ln{p_{\theta}(x,z)}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z|x)}] \ &= \mathbb{E}{z \sim q{\phi}}[\ln{p_{\theta}(x)}+\ln{p_{\theta}(z|x)}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z|x)}] \ &= \underbrace{\mathbb{E}{z \sim q{\phi}}[\ln{p_{\theta}(x)}]}{\text{□z□□□□□□□□□□□□□}}+\mathbb{E}{z \sim q_{\phi}}[{\ln{p_{\theta}(z|x)}}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z|x)}] \ &= \underbrace{\ln{p_{\theta}(x)}}{\text{□□□□□□□□□/□□}}+\underbrace{\mathbb{E}{z \sim q_{\phi}}[{\ln{p_{\theta}(z|x)}}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z|x)}]}{\text{KL□□}} \ &= \underbrace{\ell(\theta;x)}{\text{□□□□□□□□□/□□}}-\underbrace{KL(q_{\phi}(z|x)||p_{\theta}(z|x))}{\text{KL□□}} \end{align} $$ □□□□□□ $$ \underbrace{\ell(\theta;x)}{\text{□□□□□□□□□/□□}} = \underbrace{\mathcal{L}(q,\theta)}{\text{□□□□ELBO}} + \underbrace{KL(q{\phi}(z|x)||p_{\theta}(z|x))}_{\text{KL□□}} $$

□□□□□□□□□□□□□□□

$$ \begin{align} \mathcal{L}(q,\theta) &= \mathbb{E}{z \sim q{\phi}}[\ln{p_{\theta}(x,z)}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z)}] \ &= \mathbb{E}{z \sim q{\phi}}[\ln{p(z)}+\ln{p_{\theta}(x|z)}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z)}] \ &= \mathbb{E}{z \sim q{\phi}}[\ln{p(z)}]+\mathbb{E}{z \sim q{\phi}}[{\ln{p_{\theta}(x|z)}}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z)}] \ &= \mathbb{E}{z \sim q{\phi}}[{\ln{p_{\theta}(x|z)}}]+\underbrace{\mathbb{E}{z \sim q{\phi}}[{\ln{p(z)}}] - \mathbb{E}{z \sim q{\phi}}[\ln{q_{\phi}(z)}]}{\text{KL□□}} \ &= \mathbb{E}{z \sim q_{\phi}}[{\ln{p_{\theta}(x|z)}}] - \underbrace{KL(q_{\phi}(z|x)||p(z))}{q{\phi}(z)\text{□□□}p(z)\text{□KL□□}} \end{align} $$

□□□□□□□□□□ $q_{\phi}(z)$ □□ z □□□ $p_{\theta}(z|x)$ □□□□□□$ \mathcal{L}(q,\theta)$ □□□□□□□□□□□□ $\ell(\theta;x)$ □□□□□□□□□□□□ $q_{\phi}(z) = p_{\theta}(z|x)$□ □□□□□□□□□□$ q_{\phi}(z)=q_{\phi}(z|x)$□□□ $$ \begin{align} \mathcal{L}(q,\theta) &= \underbrace{\mathbb{E}{z \sim *q{\phi}(z|x)}[{\ln{p_{\theta}(x|z)}}]}{\text{□□□(reconstruction term)}}- \underbrace{KL(q{\phi}(z|x)||p(z))}_{\text{□□□□□(prior matching term)}} \ &= \ell(\theta;x) \end{align} $$*

□□□□

□□	□□	□□□□	□□□□
$p(z)$	□□□□□□□□	□□□□□□□*N(0,I)*	□
$q_{\phi}(z	*x)$*	□□□□□□	□□□□□□*ϕ*
$p_{\theta}(x	*z)$*	□□□□□□	□□□□□□*θ*
$p_{\theta}(z	*x)$*	□□□□□□	□□□□□□□□□□□□
$p_{\theta}(x)$	□□□□□□□	□□□□□□□□□□□□	□□*VAE*

- *$p(z)$*□□□□□□□□□*z*□□□□□□□□□□□□□□□
- *$p(z|x)$*□□□□□□□*z*□□□□□□□□□
- □□□□□□□□□□□□□□□□□□*z*□□□□□□□□□□*z*□□□□□□□□□□□□□□□□□□□□□□□□□□□*z*□□□□□□□□□
- □□□□□□□□□□□□□□□□□□□*z*□□□□□□□□*ELBO*□□□□□□□□□□□□

□□□□-□□□

□□□□*$q_{\phi}(z|x)$*□□□□□□□□□□□□□□□□□□□□□□□□□□□□ *μ_z* □ *Σ_z* □□□□□□□□□□□□□□□□□ □□□ *$q_{\phi}(z|x) = \mathcal{N}(\mu_z,\Sigma_z)$*□□□□□□□ *$Z $*□□□□□□□□□□□□ □□□□□□ *$\Sigma_z$* □□□□□□□□□□□□□□□□□ *0* □□□□□□□□□□□□

□ *VAE* □□*Z* □□□ □□□□□□□□*X*□□□□□*Z* □□□□□□□*X*□□□□□*$Z $*□□□□□□□□□□□□□□*$Z$*□□□□□□□□ *$\mu_z$*□□□□□ *$\Sigma_z$* □□*$x$*□□□□□□ *$$ \begin{align} \mu_z &= \mu_{\phi}(x) = encoder_{\phi}(x) \ \Sigma_z &= \Sigma_{\phi}(x) = encoder_{\phi}(x) \end{align} $$*

***KL*□□-□□□**

$$ \begin{align} KL(q_{\phi}(z)||p(z)) &= KL(\mathcal{N}(\mu_z,\Sigma_z)||\mathcal{N}(0,I)) \ &= \frac{1}{2}(tr(\Sigma_z)+\mu_z^T\mu_z-k-\log{det(\Sigma_z)}) \end{align} $$

□□□□□□□□□□□□□□□□*k*□□□□□□□□□□□□□□□□□□□□□□□*$q_{\phi}(z|x)$* □□□□*z*□□□

□□□□-□□□

$\underbrace{\mathbb{E}{z \sim q{\phi}(z|x)}[{\ln{p_{\theta}(x|z)}}]}{\text{□□□(reconstruction term)}}$□□□□□□□□□□□□*$Z$*□□□*$X$*□*$z$*□*$x$*□□□□□□□*$z$*□*$X$*□□□□□*$\mu_x$*□□□□□□□□□□□□□□□□□ *$z$*□*$X$*□□□□*$\Sigma_x$*□□□□□□ □□□□□□□□□□□□*$X$* □□□□□□□□□□□*$I$* *$$ \mu_x = \mu{\theta}(z) = decoder_{\theta}(z) $$* □□□□□□□□*Z*□□ □□□□□□□□□□□□*$q_{\phi}(z|x)$*□□□□□□□*z*□□□□□□□□□□ *$$ \mathbb{E}{z \sim q{\phi}(z|x)}[{\ln{p_{\theta}(x|z)}}] \approx \frac{1}{L}\sum_{l=1}^L[\ln{p_{\theta}(x|z^{(l)})}] $$* □□□□ *L*□□□□□□□□□□□□□□□□□□□□□□□□*$L=1$*□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□*z*□□□□□□ □□□□□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□*VAE* □□□□ □□□□□□□□□*reparameterization trick*□□□□□□□□ □□□□

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□$q_{\phi}(x|z)$□□□□□z□□□□□□□□□□□□□□□$\mathcal{N}(\mu_z,\Sigma_z)$□□□□□□□□□□□□□ □□□**$\mathcal{N}(0,I)$**□□□□□□□□

□□ϵ□□□□□□□□□□□□□□□$\epsilon \sim \mathcal{N}(0,I)$□□□□□□□□□□□$\mathcal{N}
(\mu_z,\Sigma_z)$□□□□□□□□□□□□□□□ $$ \begin{align} z &= \mu_z + \sqrt{\Sigma_z} \odot \epsilon \
&=\mu_{\theta}(x)+\sqrt{\Sigma_{\phi}(x)} \odot \epsilon,\epsilon \sim \mathcal{N}(0,I) \end{align}
$$ □□□□□□□□□□□□□□□□$\mathcal{N}(0,I)$□□□□□□□□□□□□□□□□□□□□□z□□□□□$\odot$□□□□□□□□□□□□□□
encoder□□□$\mu_{\phi}(x)$□□□□□□□□□□□□□□□encoder□□□□□□$\sqrt{\Sigma_{\phi}(x)}$□□□□□□□□□□□□
ϵ □□□□□□□□□□□□□□

□□□□□□

□□□□□□□□□□□□□□□ $$ p(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}exp{-\frac{1}{2}(x-\mu)^T\Sigma
^{-1}(x-\mu)} $$ □□□□□□□$p_{\theta}(x|z)$□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□$p_{\theta}(x|z)$□□□□ $$
\begin{align} p(x) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}exp{-\frac{1}{2}(x-\mu)^T\Sigma
^{-1}(x-\mu)} \ & \propto exp{-\frac{1}{2}(x-\mu_x)^T(x-\mu_x)} \ &= exp{-\frac{1}{2}(x-
\mu_{\theta}(z))^T(x-\mu_{\theta}(z))} \ &= exp{-\frac{1}{2}(x-decoder(z))^T(x-x-decoder(z))} \
\end{align} $$ □□□□□□$ELBO$□□□□ $$ \begin{align} \mathcal{L}(q,\theta) &=
\underbrace{\mathbb{E}*{z \sim q*{\phi}}[{\ln{p_{\theta}(x|z)}}]}*{□□□□□□} -
\underbrace{KL(q*{\phi}(z|x)||p(z))}*{□□□□□□} \ &= \frac{1}{L}\sum*{l=1}^L[\ln{p_{\theta}
(x|z^{(l)})}] - KL(\mathcal{N}(\mu_z,\Sigma_z)||\mathcal{N}(0,I)) \ &\propto \frac{1}
{L}\sum_{l=1}^L[-\frac{1}{2}(x-\mu_x)^T(x-\mu_x)] - [\frac{1}{2}(tr(\Sigma_z)+\mu_z^T\mu_z-k-
\log{det(\Sigma_z)})]\ & \propto -\frac{1}{L}\sum_{l=1}^L[(x-\mu_x)^T(x-\mu_x)] -
[(tr(\Sigma_z)+\mu_z^T\mu_z-k-\log{det(\Sigma_z)})]\ \end{align} $$ □□□ $$ \begin{align} \mu_x &=
\mu_{\theta}(z^{(l)}) = decoder(z^{(l)}) \ z^{(l)} &= \mu_z + \sqrt{\Sigma_z} \odot \epsilon,
\epsilon \sim \mathcal{N}(0,I)\ \mu_z &= \mu_{\phi}(x) = encoder(x)*{\mu_z}\ \Sigma_z &=
\Sigma*{\phi}(x) = encoder(x)*{\Sigma_z} \end{align} $$ □□□□□□□□□□□$\mathcal{L}(q,\theta) $□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ $$ \frac{1}{L}\sum*{l=1}^L[\underbrace{(x-\mu_x)^T(x-
\mu_x)}_{□□□□}] $$ □ $$ (tr(\Sigma_z)+\mu_z^T\mu_z-k-\log{det(\Sigma_z)}) $$ □□□□□□□□□□□□□□□□□□
□□□□□

□□□□

- □□□□□□□□□□□z□□□x
- □□□□□□□□□□□□□□□z□□□□□□□□□□□□□□□□
  - □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
- □□□□□x□□x□□□
  - □□□

$$ \mathcal{L}(q,\theta) = \underbrace{\mathbb{E}*{z \sim q*{\phi}(z|x)}[{\ln{p_{\theta}(x|z)}}]}
{\text{□□□(reconstruction term)}}- \underbrace{KL(q{\phi}(z|x)||p(z))}_{\text{□□□□□(prior matching
term)}} $$

□□□□

1. □□□□x
2. □□□□□□□□□□□□□□μ,σ
3. □□□□□□□z
4. □□□□z□□□□ $\hat{x}$
5. □□ x □ $\hat{x}$□□□□□□□
6. □□□□□□□□□$\mathcal{N}(0,1)$□ KL □□
7. □□□□□□□□□□□□□□□□□□□

AE□Auto-encoder □□□□□

□□□□□

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

1. [illegible]**Encoder**[illegible]
    - [illegible]"[illegible]"[illegible]"[illegible]"[illegible]
    - [illegible]
    - [illegible]
2. [illegible]**Latent Space**[illegible]
    - [illegible]
    - [illegible]
    - [illegible]
3. [illegible]**Decoder**[illegible]
    - [illegible]
    - [illegible]
    - [illegible]

[illegible]

1. [illegible]
2. [illegible]784[illegible]32[illegible]
3. [illegible]
4. [illegible]"[illegible]"[illegible]
5. [illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]