Introduction to Variational Autoencoders (VAEs)

- **Goal of VAEs**: Learn latent representations of data for generation
- **Components**:
  - **Encoder**: Maps data x to latent variable z
  - **Decoder**: Maps latent variable z back to data x
- **Objective**: Maximize the likelihood of observed data through latent representations

Why Use VAEs?

- **Data Complexity**: Real-world data has variability
- **Latent Structure**: Captures underlying patterns in data
- **Probabilistic Framework**: Encodes uncertainty in latent space
- **Use Cases**:
    - Image generation
    - Data compression
    - Anomaly detection

Probabilistic Modeling: The Role of $p(x|z)$

- $p(x|z)$: Models the probability of data x given latent variable z
- **Purpose**:
  - Captures variability in data (e.g., handwriting styles)
  - Allows probabilistic generation of x from latent space z
- **No 1:1 Mapping**: $p(x|z)$ accommodates real-world complexity

Introduction to Information Theory

- **Information**: Quantifies "surprise" of an event
- **Formula**: $I(x) = -\log p(x)$
  - High surprise for rare events
  - Low surprise for common events
- **Example**:
  - Rare event: High $I(x)$, large $-\log p(x)$
  - Common event: Low $I(x)$, small $-\log p(x)$

What is KL Divergence?

- **Definition**: Measures difference between two distributions
- **Formula**:

$$D_{KL}(q(z) \| p(z)) = \mathbb{E}_{q(z)} \left[\log \frac{q(z)}{p(z)} \right]$$

- **Purpose**: Quantifies inefficiency of approximating $p(z)$ with $q(z)$

KL Divergence: Formula Breakdown

1. **Expanded Form**:

$$D_{KL}(q(z)\|p(z)) = \int q(z) \log \frac{q(z)}{p(z)} \, dz$$

2. **Interpretation**:
    - **Non-negative**: $D_{KL} \geq 0$
    - **Zero**: Only if $q(z) = p(z)$ exactly
3. **Use in VAEs**:
    - Keeps $q(z|x)$ close to $p(z)$

Marginal Likelihood in VAEs: $p(x)$

- **Goal**: Maximize likelihood of data x
- **Formula**:

$$p(x) = \int p(x|z)p(z)\, dz$$

- **Challenge**: Intractable integral over z in high dimensions

Using $q(z|x)$ to Approximate $p(x)$

- **Rewrite** $\log p(x)$ **with** $q(z|x)$:

$$\log p(x) = \log \int \frac{p(x|z)p(z)}{q(z|x)} q(z|x)\, dz$$

- $q(z|x)$: Serves as an approximation to $p(z|x)$

Applying Jensen's Inequality to $\log p(x)$

- **Jensen's Inequality**: For concave $\log$

$$\log \mathbb{E}[X] \geq \mathbb{E}[\log X]$$

- **Application**:

$$\log p(x) \geq \int q(z|x) \log \frac{p(x|z)p(z)}{q(z|x)} \, dz$$

- **Result**: A lower bound on $\log p(x)$

Deriving the Evidence Lower Bound (ELBO)

- **ELBO Definition**:

$$\mathcal{L}_{\text{ELBO}} = \int q(z|x) \log \frac{p(x|z)p(z)}{q(z|x)} \, dz$$

- **Purpose**: Provides a bound we can maximize instead of $\log p(x)$

Expanding the ELBO Formula

- **Breakdown**:

$$\mathcal{L}_{\text{ELBO}} = \int q(z|x) \left(\log p(x|z) + \log p(z) - \log q(z|x) \right) dz$$

1. **Reconstruction**: $\mathbb{E}_{q(z|x)}[\log p(x|z)]$
2. **KL Divergence**: $-D_{KL}(q(z|x) \| p(z))$

ELBO Interpretation: Balancing Objectives

- **Reconstruction Term**: Maximizes data reconstruction accuracy
  - $\mathbb{E}_{q(z|x)}[\log p(x|z)]$
- **KL Divergence Term**: Regularizes latent space
  - $-D_{KL}(q(z|x) | p(z))$
- **Goal**: Balance reconstruction with structured latent space

Summary of Key Concepts in VAEs

- **VAEs**: Learn probabilistic latent representations for data generation
- **KL Divergence**: Measures the difference between $q(z|x)$ and $p(z)$
- **ELBO**: Lower bound on $\log p(x)$, combines reconstruction and regularization