Limitations of Standard Autoencoders

- They struggle to learn meaningful latent spaces.
- Lack of control over latent representations.
- Enter **Variational Autoencoders (VAEs)**.

Variational Autoencoders (VAEs)

- **VAEs** extend autoencoders to probabilistic models.
- Instead of a fixed latent (z), VAEs learn a probability distribution (p(z|x)).
- Outputs are sampled from this distribution.

VAE Architecture

```
Input --> Encoder --> Latent Distribution --> Decoder --> Output
```

- **Encoder**: Produces (\mu) and (\sigma^2) for latent distribution.
- **Latent Distribution**: Sampled using reparameterization trick.
- **Decoder**: Reconstructs (x) from samples (z).

Latent Space in VAEs

- In VAEs, the latent space represents a distribution.
- Assumes $z \sim \mathcal{N}(\mu, \sigma^2)$, a Gaussian distribution.

VAE Loss Function

The VAE loss combines two terms:

1. **Reconstruction Loss**:

$$L_{\text{recon}}(x, \hat{x}) = ||x - \hat{x}||^2$$

2. **KL Divergence** (regularizer):

$$L_{\text{KL}} = D_{\text{KL}}(q(z|x)||p(z))$$

KL Divergence in VAEs

The KL divergence regularizes the latent space:

$$D_{\mathrm{KL}}(q(z|x)||p(z)) = \int q(z|x)\log\left(\frac{q(z|x)}{p(z)}\right)dz$$

- Encourages (q(z|x)) to be close to the prior (p(z)).

Reparameterization Trick

To allow backpropagation through the sampling process:

- Replace $z \sim \mathcal{N}(\mu, \sigma^2)$ with:

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

Kullback-Leibler (KL) Divergence

The formula for KL divergence between a reference probability distribution P and a second probability distribution Q is:

$$D_{KL}(P||Q) = \sum_{x \in \chi} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

- **Interpretation**: The expected excess surprise from using Q as a model instead of P.

Reconstruction Loss

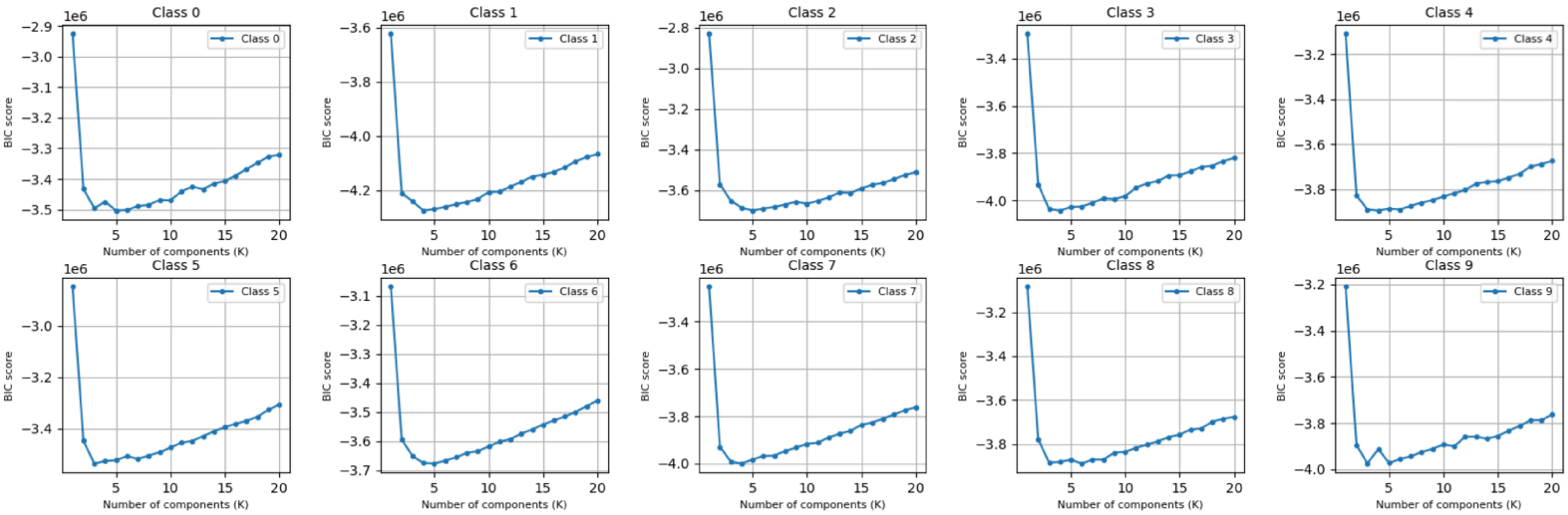
The code to compute the reconstruction loss:

```python
from torch.nn import functional as F

def reconstruction_loss(x, x_hat):
    return F.binary_cross_entropy(x_hat, x, reduction="sum")
```

- Inputs: `x` and `x_hat` have dimensions $(N, 1, H, W)$.
- **Explanation**: This computes the negative log-likelihood of the Bernoulli distribution.

BIC Curve



Class 0
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9
1e6
BIC score
Number of components (K)

BIC Loss

$$\frac{1}{2}$$