

# AUTO-ENCODING VARIATIONAL BAYES

**AUTHORS:**  
DIEDERIK P. KINGMA, MAX  
WELLING

|  
**GROUP 8**  
RACHITHA CHANDRA BHANU  
THOTA BHUVANA CHANDRA  
SHEHARIYAR FIRDOUS SHAIKH  
SAI SURYA MADA

# THE PROBLEM

- Efficient inference and learning in probabilistic models with:

Intractable posterior distributions

Large datasets

Continuous latent variables

- Traditional variational inference requires model-specific derivations
- Monte Carlo methods are often too slow for large-scale data

# KEY CONTRIBUTIONS

- The Variational Autoencoder (VAE) framework
- The reparameterization trick for backpropagation through stochastic nodes
- Stochastic Gradient Variational Bayes (SGVB) estimator
- Auto-Encoding Variational Bayes (AEVB) algorithm
- Scalable inference for continuous latent variable models

# VARIATIONAL INFERENCE BACKGROUND

- Goal: Approximate intractable posterior  $p(z|x)$  with tractable  $q(z|x)$
- Minimize KL divergence:  $\text{KL}(q(z|x) \parallel p(z|x))$
- Equivalent to maximizing the Evidence Lower Bound (ELBO)
- $\text{ELBO} = \mathbb{E}_{q\phi(z|x)}[\log p\theta(x|z)] - \text{KL}(q\phi(z|x) \parallel p(z))$
- Traditional methods: mean-field approximations, require conjugacy

# THE REPARAMETERIZATION TRICK

- Problem: Can't backpropagate through sampling operation
- Solution: Reparameterize random variable  $z$
- Instead of:  $z \sim q(z|x)$
- Write as:  $z = g(\varepsilon, x)$  where  $\varepsilon \sim p(\varepsilon)$
- Example:  $z \sim N(\mu, \sigma^2)$  becomes  $z = \mu + \sigma \cdot \varepsilon$  where  $\varepsilon \sim N(0,1)$
- Now gradients can flow through  $\mu$  and  $\sigma$ .

# VAE ARCHITECTURE

**Encoder (Recognition Model):  $q=\varphi(z|x)$**

- Neural network that maps input  $x$  to latent distribution parameters
- Outputs mean  $\mu$  and variance  $\sigma^2$  of approximate posterior

**Decoder (Generative Model):  $p=\theta(x|z)$**

- Neural network that maps latent  $z$  to data distribution parameters
- Reconstructs input from latent representation
- Both trained jointly to maximize ELBO

# THE LOWER BOUND DECOMPOSITION

$$\mathcal{L}(\theta, \phi; \mathbf{x}^{(i)}) = -D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) || p_\theta(\mathbf{z})) + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x}^{(i)})} [\log p_\theta(\mathbf{x}^{(i)}|\mathbf{z})]$$

- First term: Regularization (KL divergence between approximate and prior)
- Second term: Reconstruction loss (negative expected log-likelihood)
- Trade-off between reconstruction quality and latent space regularity

# SGVB ESTIMATOR

- Stochastic Gradient Variational Bayes
- Unbiased gradient estimator
- Can be computed efficiently with automatic differentiation

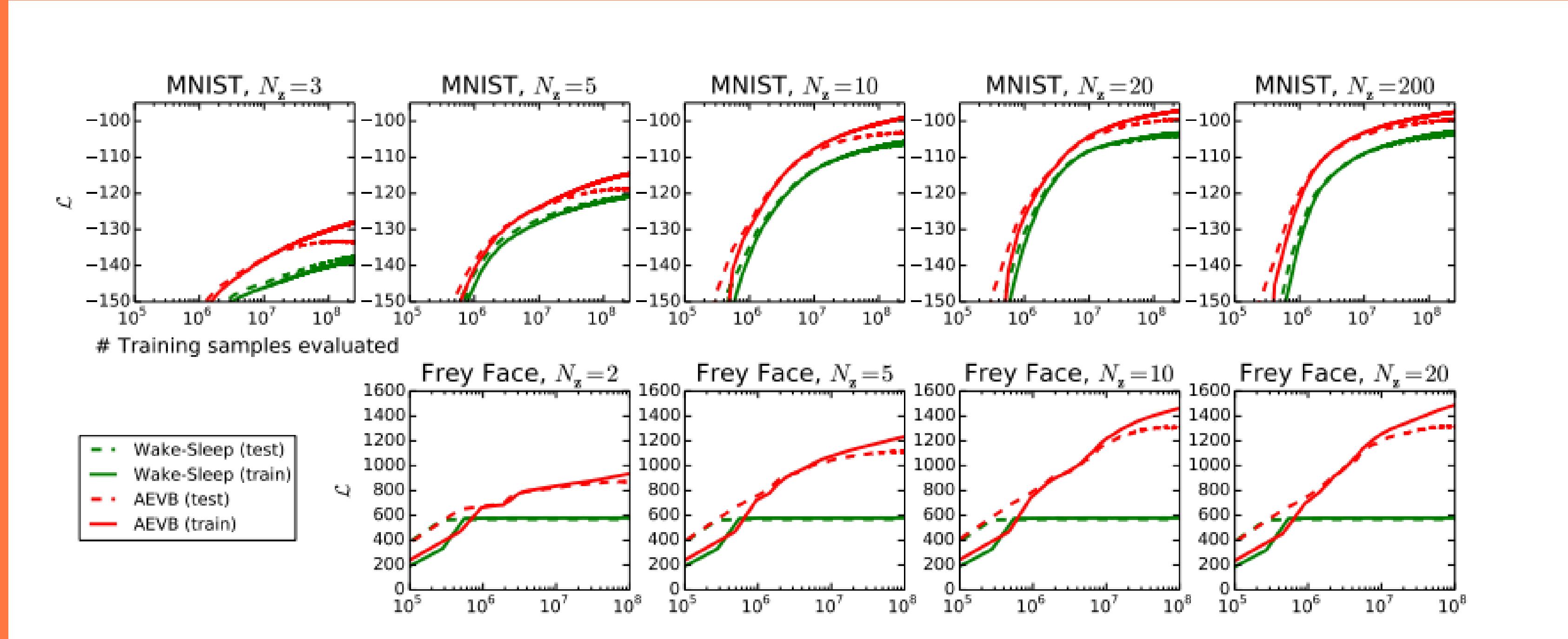
$$\tilde{\mathcal{L}}^A(\theta, \phi; \mathbf{x}^{(i)}) = \frac{1}{L} \sum_{l=1}^L \log p_\theta(\mathbf{x}^{(i)}, \mathbf{z}^{(i,l)}) - \log q_\phi(\mathbf{z}^{(i,l)} | \mathbf{x}^{(i)})$$

where  $\mathbf{z}^{(i,l)} = g_\phi(\epsilon^{(i,l)}, \mathbf{x}^{(i)})$  and  $\epsilon^{(l)} \sim p(\epsilon)$

# AEVB ALGORITHM

- Initialize parameters  $\theta, \varphi$
- For each iteration:
  - Sample minibatch of datapoints
  - Sample  $\varepsilon$  from  $p(\varepsilon)$  for each datapoint
  - Compute gradients using SGVB estimator
  - Update  $\theta, \varphi$  using gradient.
  - Works with any continuous latent variable model

# EXPERIMENTS





(a) Learned Frey Face manifold

6 6 6 6 6 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
9 4 4 4 2 2 2 2 2 0 0 0 0 0 0 0 0 0 0 0 0  
9 2 2 2 2 2 2 2 2 6 6 6 6 0 0 0 0 0 0 0 2  
9 9 2 2 2 2 2 2 3 3 5 5 5 5 0 0 0 0 0 2  
9 9 2 2 2 2 2 2 3 3 3 3 5 5 5 5 5 5 5 3  
9 9 9 2 2 2 3 3 3 3 3 3 5 5 5 5 5 3  
9 9 9 9 2 2 3 3 3 3 3 3 5 5 5 5 5 3  
9 9 9 9 9 8 3 3 3 3 3 3 3 5 5 5 5 3  
9 9 9 9 9 8 3 3 3 3 3 3 3 8 8 8 8 8 7  
9 9 9 9 9 8 3 3 3 3 3 3 3 8 8 8 8 8 7  
7 9 9 9 9 8 8 8 8 8 8 8 8 8 8 8 8 8 7  
7 9 9 9 9 8 8 8 8 8 8 8 8 8 8 8 8 8 7  
7 9 9 9 9 8 8 8 8 8 8 8 8 8 8 8 8 8 7  
7 9 9 9 9 8 8 8 8 8 8 8 8 8 8 8 8 5 7  
7 9 9 9 9 8 8 8 8 8 8 8 8 8 6 6 6 6 5 7  
7 9 9 9 9 9 8 8 8 8 8 8 8 6 6 6 6 6 5 7  
7 9 9 9 9 9 9 8 8 8 8 8 6 6 6 6 6 6 5 7  
7 9 9 9 9 9 9 9 8 8 8 8 6 6 6 6 6 6 6 5 7  
7 9 9 9 9 9 9 9 9 8 8 8 6 6 6 6 6 6 6 6 1  
7 9 9 9 9 9 9 9 9 8 8 8 6 6 6 6 6 6 6 6 1  
7 9 9 9 9 9 9 9 9 8 8 8 6 6 6 6 6 6 6 6 1  
7 9 9 9 9 9 9 9 9 8 8 8 6 6 6 6 6 6 6 6 1  
7 7 7 7 7 7 7 7 7 1 1 1 1 1 1 1 1 1 1 1 1

(b) Learned MNIST manifold

8 6 1 7 8 1 4 8 2 8 5 1 6 5 7 0 7 6 7 2 1 8 7 1 3 8 5 7 3 8 1 2 0 8 7 2 1 9 0 0  
9 6 8 5 9 6 0 3 1 9 8 5 9 4 6 3 2 1 6 2 2 3 8 2 7 9 3 5 3 8 7 5 1 9 1 1 7 1 4 4  
1 1 1 1 3 6 9 1 7 9 6 1 5 2 2 8 8 4 3 3 3 5 9 9 4 3 9 5 1 6 8 9 6 2 0 3 2 9 2 9  
8 9 0 8 6 9 1 4 6 3 2 1 6 8 9 1 0 0 4 1 1 9 1 8 8 3 4 9 7 1 9 8 4 3 1 7 0 6 1  
8 2 3 3 3 1 3 3 6 5 1 9 2 0 1 5 3 5 9 4 7 3 6 4 2 0 2 6 5 5 4 7 9 1 9 7 9 1 5  
6 9 9 8 6 1 6 6 6 6 6 6 6 1 4 9 1 7 5 8 5 7 7 0 5 2 2 7 4 5 6 2 2 4 2 8 6 2 8 1  
9 5 2 6 6 5 1 8 9 9 1 3 4 3 9 1 3 2 7 0 6 7 4 3 6 2 8 5 5 2 1 5 9 2 1 6 1 3 5 3  
1 9 7 7 3 1 2 8 2 3 4 5 8 2 9 7 0 9 5 9 5 4 9 0 5 0 7 0 6 6 7 9 3 9 2 2 9 3 5 0  
0 4 6 1 2 3 2 0 8 5 6 9 9 4 2 7 2 1 2 5 7 4 1 6 2 0 3 6 0 1 4 5 2 4 3 9 0 1 5 4  
9 7 5 9 9 3 4 8 5 1 2 6 4 5 6 0 9 7 9 8 2 1 2 0 4 7 1 0 5 0 2 8 7 2 3 1 6 2 3 6

(a) 2-D latent space

(b) 5-D latent space

(c) 10-D latent space

(d) 20-D latent space

# CONCLUSION

- Auto-Encoding Variational Bayes revolutionized variational inference
- Elegant solution to intractable inference problem
- Enabled scalable learning of complex generative models
- Continues to influence modern machine learning research
- Foundational work with thousands of citations

# **THANK YOU**