Experimental Results on VAE

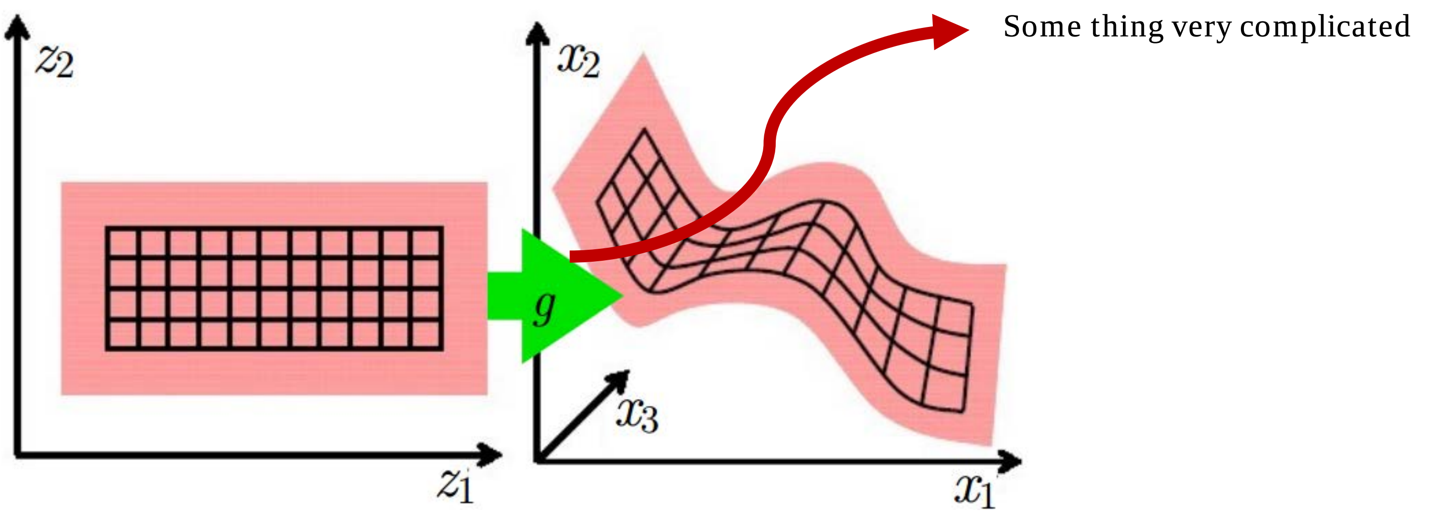
2015.11.25 산업공학특론 발표

Latent variable model

- Latent variables can extract the true explanatory factors of the (observed) original variables : generative model
- Latent variable space tends to a more simple space.

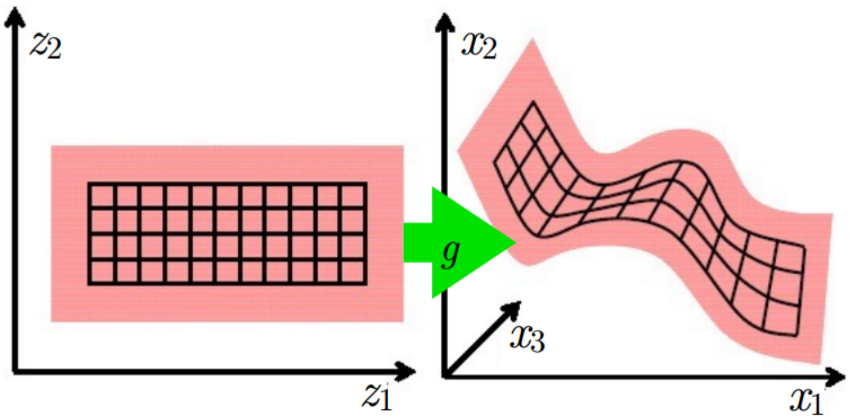
$$p(x) = \int p(x,z)\, dz \quad \text{where} \quad p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{x} \mid \boldsymbol{z})p(\boldsymbol{z})$$

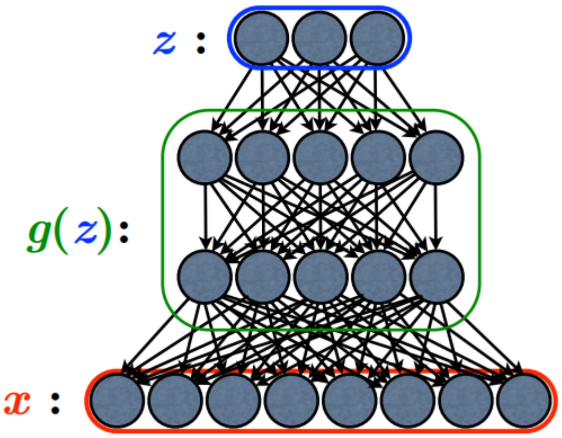
$$p(\boldsymbol{z}) = \text{something simple} \qquad p(\boldsymbol{x} \mid \boldsymbol{z}) = g(\boldsymbol{z})$$



Latent variable model

- Use neural networks as the (generative) transformation g from the latent space to the original feature space.
- For both training and inference, latent variable z must be inferred.
- This is a generative model : inferring the latent variable is hard
- The posterior $p_\theta(z|x)$ is intractable





Variational inference

- How to infer the latent variable z
- Use $q_\varphi(z|x)$ as an alternative to $P_\theta(z|x)$
- Maximize the lower bound of the likelihood $p_\theta(x) \geq L(\theta, \varphi, x)$

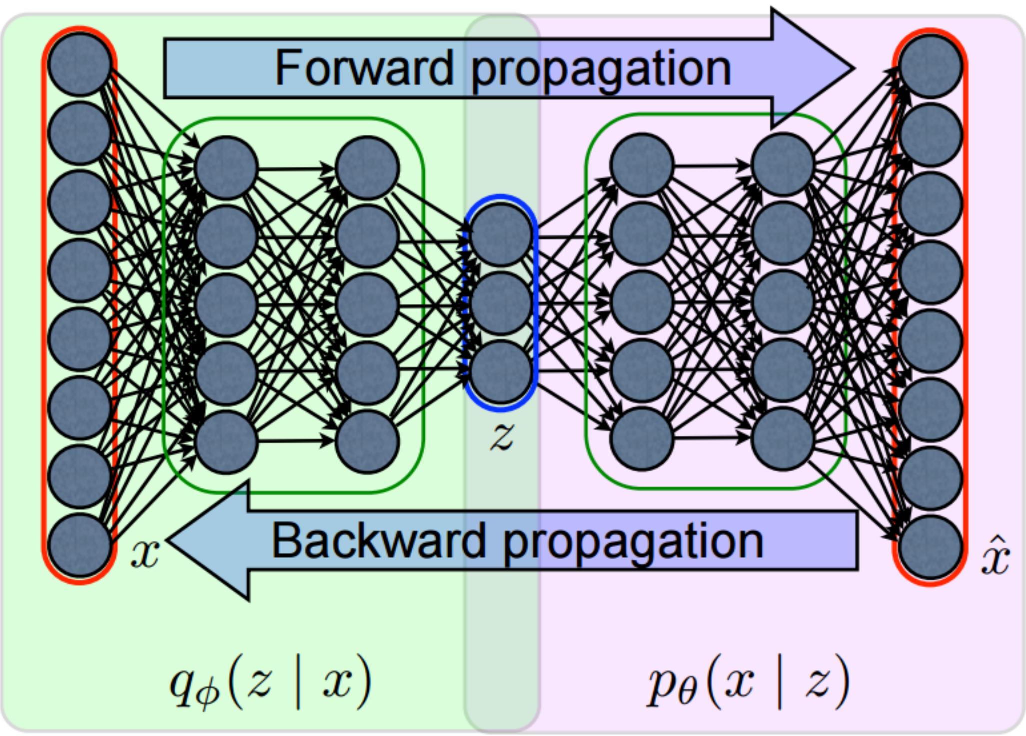
$$\begin{aligned}\mathcal{L}(\theta, \phi, x) &= \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x, z) - \log q_\phi(z \mid x)\right] \\ &= \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x \mid z) + \log p_\theta(z) - \log q_\phi(z \mid x)\right] \\ &= -D_{\mathrm{KL}}\left(q_\phi(z \mid x) \| p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x \mid z)\right]\end{aligned}$$

regularization ***term*** *reconstruction* ***term***

- Train this by using backpropgation

Variational Autoencoder(VAE)

Objective function: $\mathcal{L}(\theta, \phi, x) = -D_{\mathrm{KL}}\left(q_\phi(z \mid x) \| p_\theta(z)\right) + \mathbb{E}_{q_\phi(z \mid x)}\left[\log p_\theta(x \mid z)\right]$

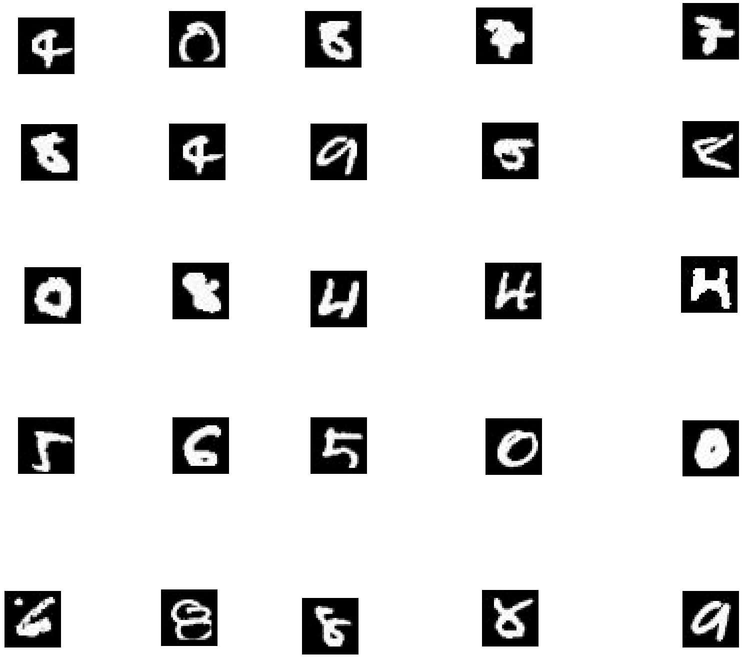


Experimental Setup

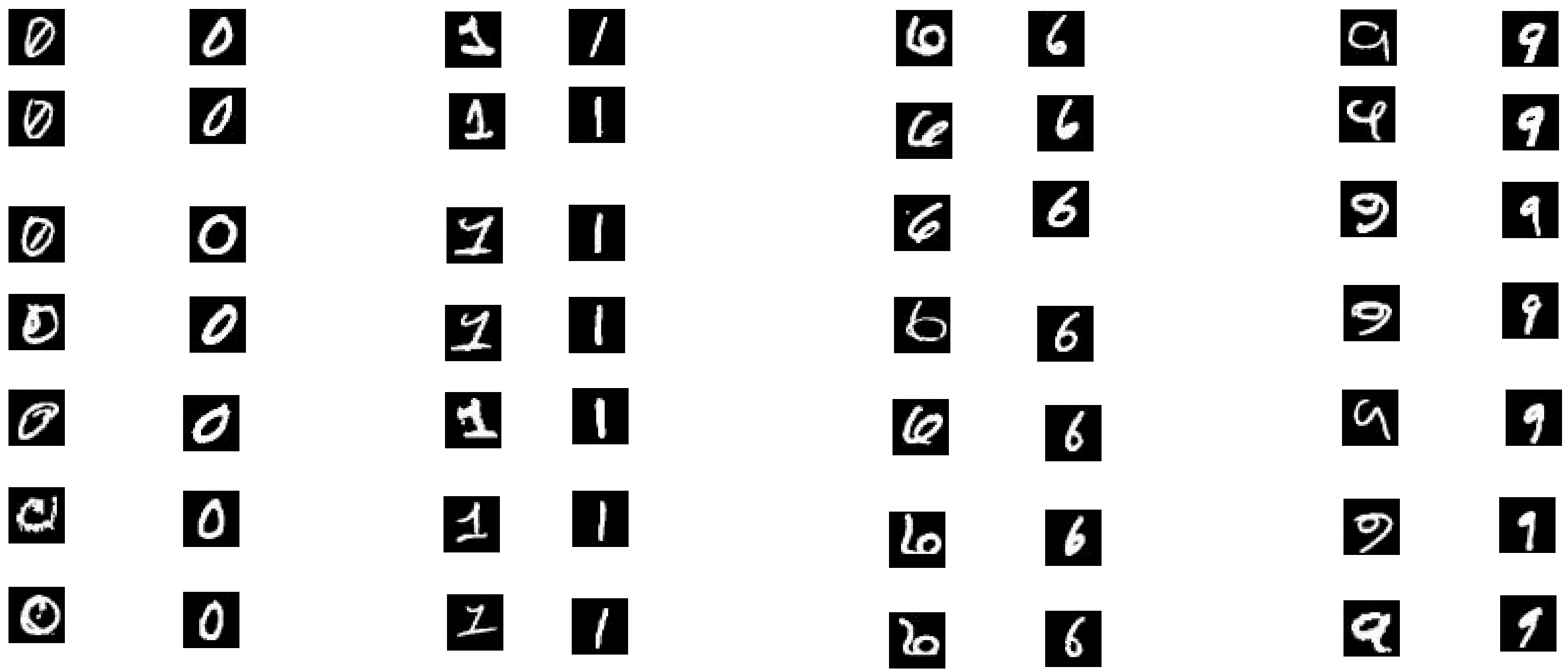
- VAE : Both the encoder and decoder has one hidden layer of size 400
- The dimension of latent variables were varied from 3, 5, and 200.
- Data : MNIST
  - Training : 50000
  - Test : 10000

- Batch size : 512
- Epochs : 200

Data points according to likelihood probability (decoder) :global

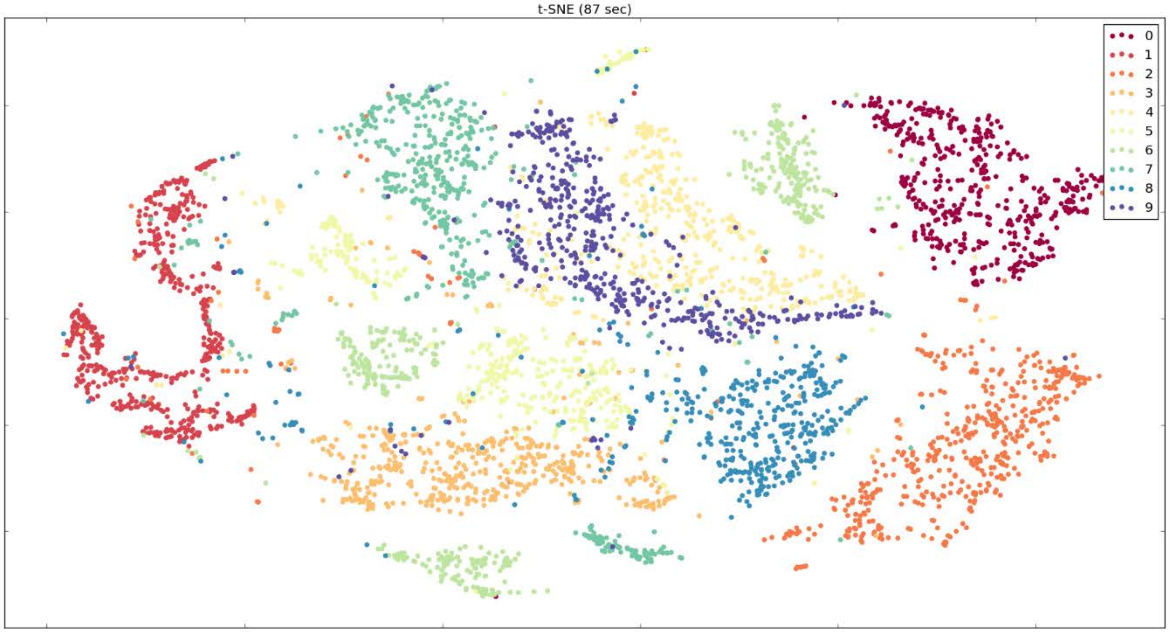


Data points according to likelihood probability (decoder) : conditional



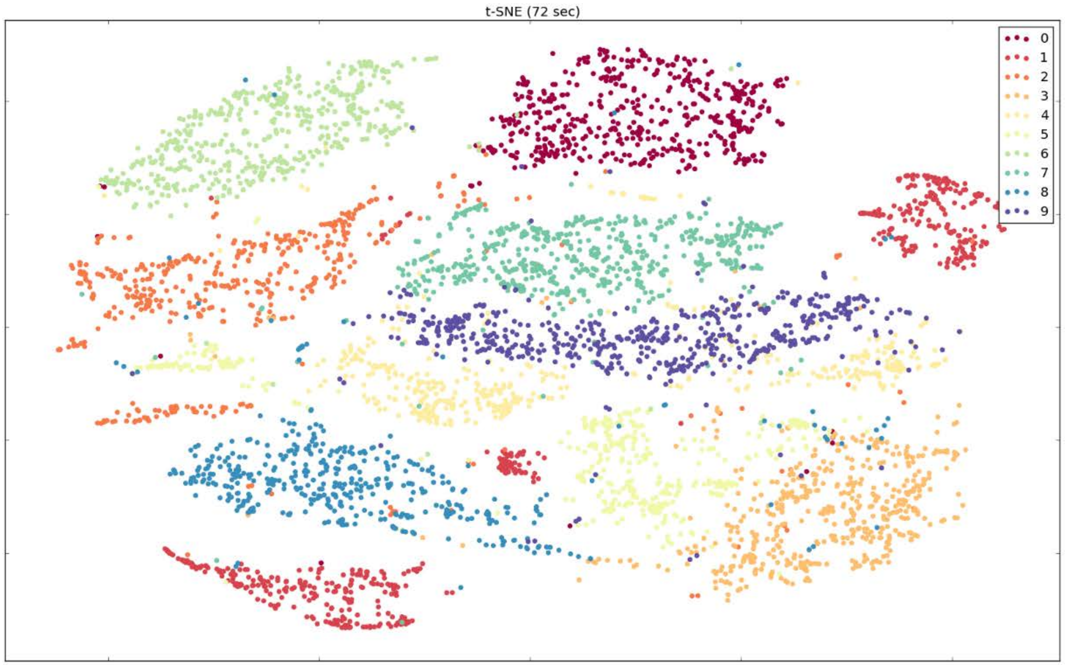
t-sne projection of latent variables

- Latent dimension 3



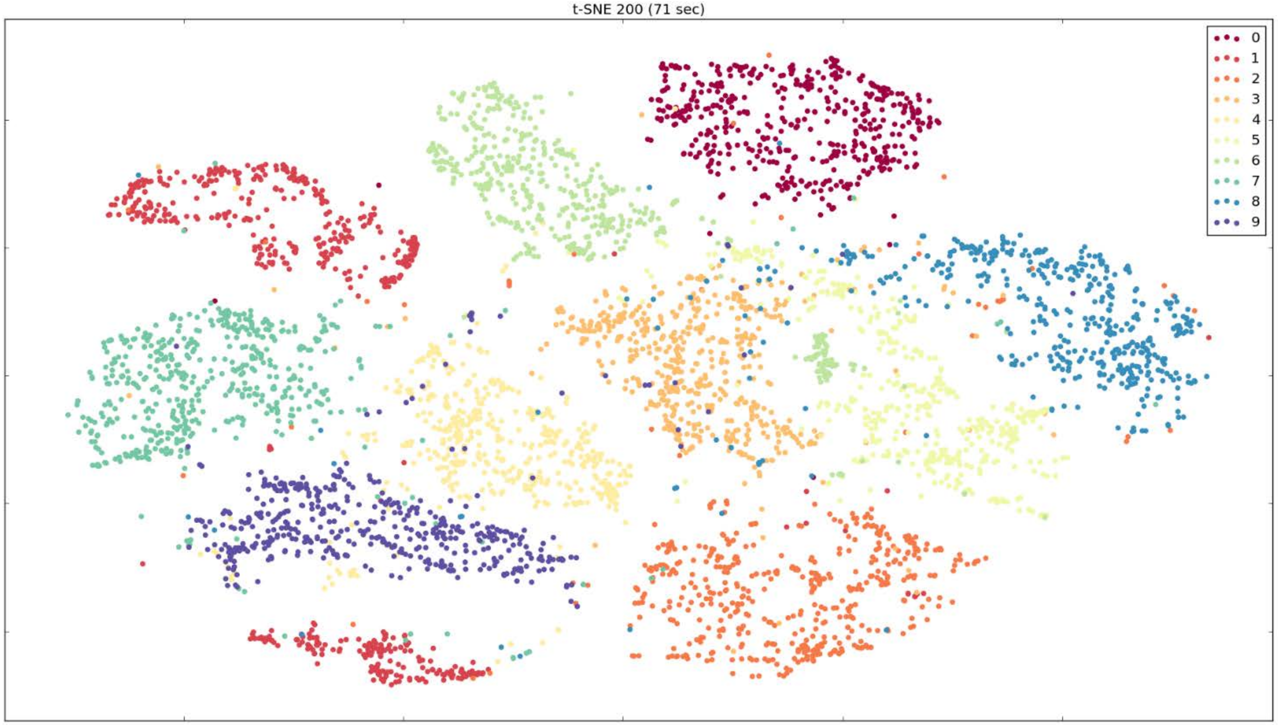
t-sne projection of latent variables

- Latent dimension 5



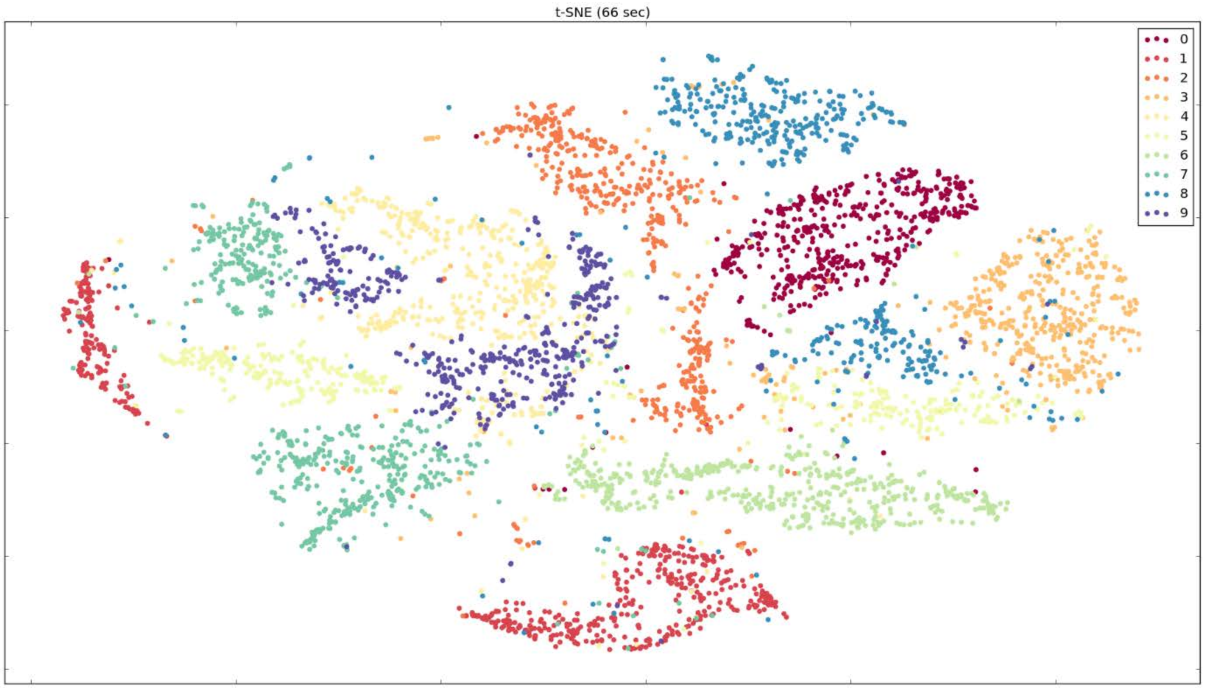
t-sne projection of latent variables

- Latent dimension 200



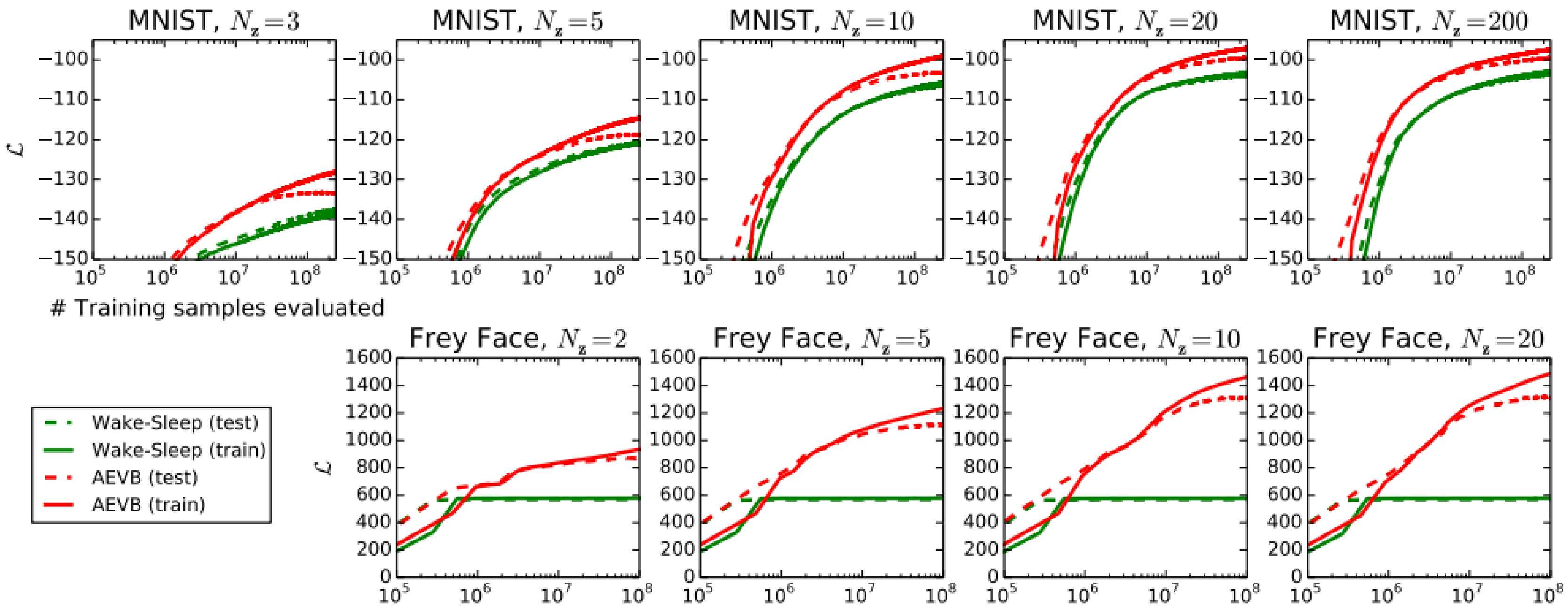
t-sne projection

- Original



Limitations

- More training epochs are required



Todo

- Run unsupervised learning (1-class svm) with using latent variables of various dimensions
  - Compare with original data
- Test with artificial data that has particular types of anomalies
  - Make connections with latent variables and anomaly type
- Combine probabilities and latent variables to find true anomalies.
- Apply semi-supervised VAE.
  - With a handful of labels on anomalies, can semi-supervised learning work?

Referenences

- Kingma, Diederik P., et al. "Semi-supervised learning with deep generative models." *Advances in Neural Information Processing Systems*. 2014.
- Kingma, Diederik P., and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114* (2013).