Practice 10

Molecular generative model (2)

Seongok Ryu

Department of Chemistry, KAIST

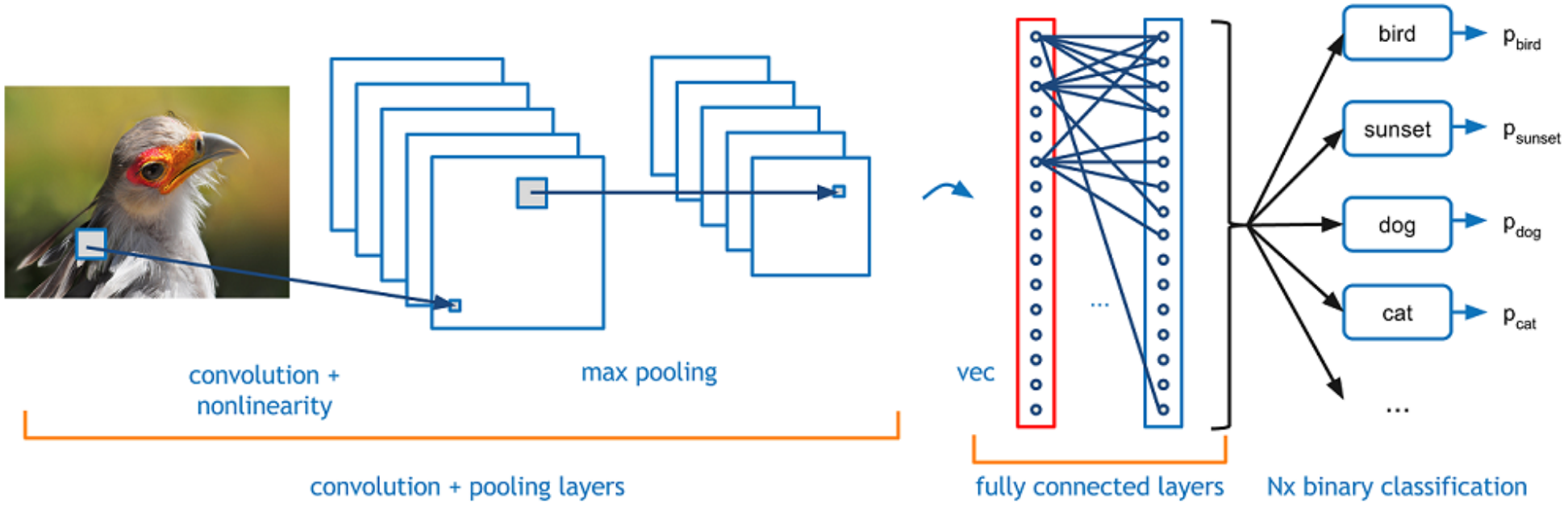
Why does deep learning work so well?

Sufficient statistics

Information distillation (Feature extraction)

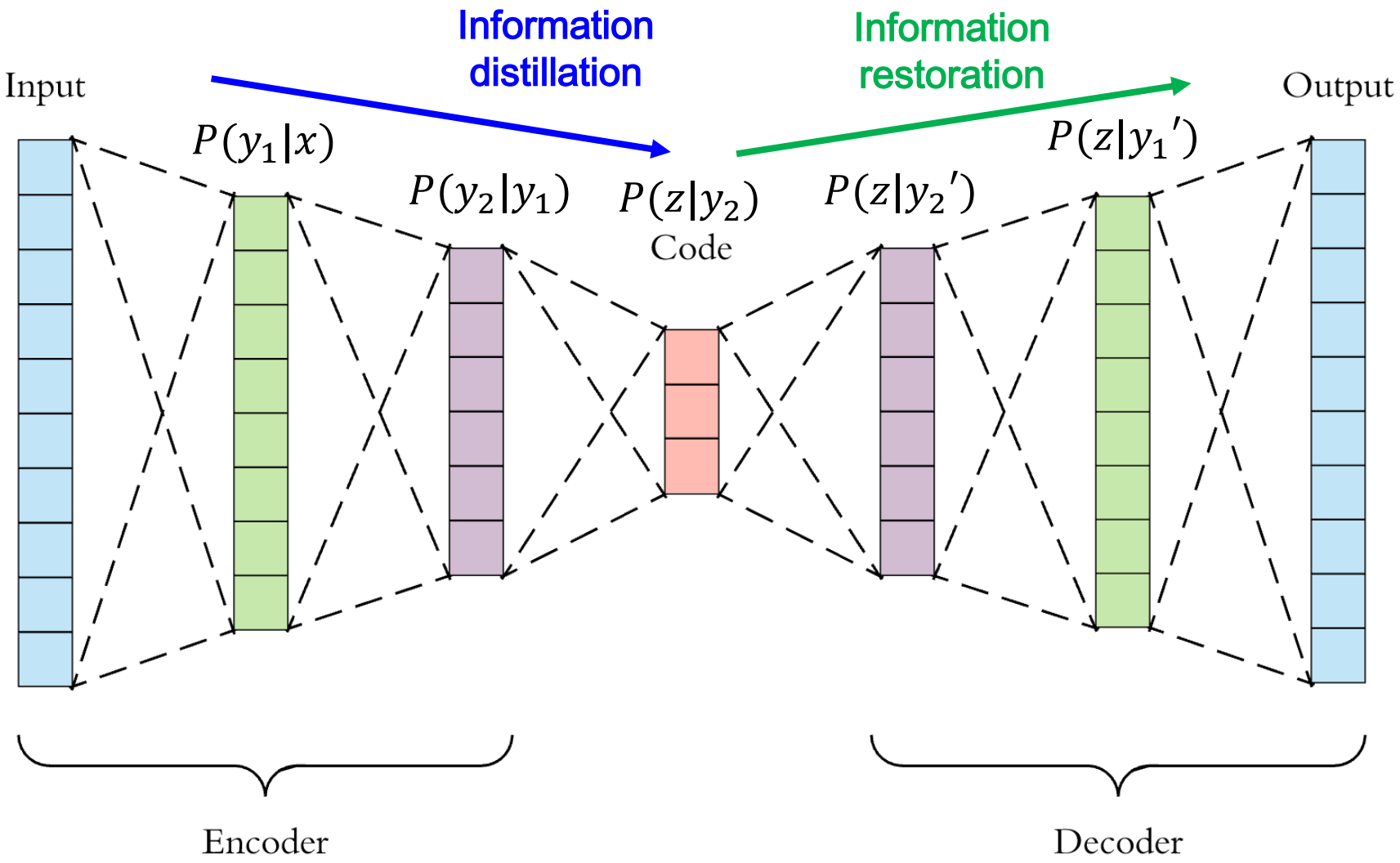
Classification based on the feature

$$x \rightarrow P(y_1|x) \rightarrow P(y_2|y_1) \rightarrow P(y_3|y_2) \rightarrow P(c = bird|y_3)$$



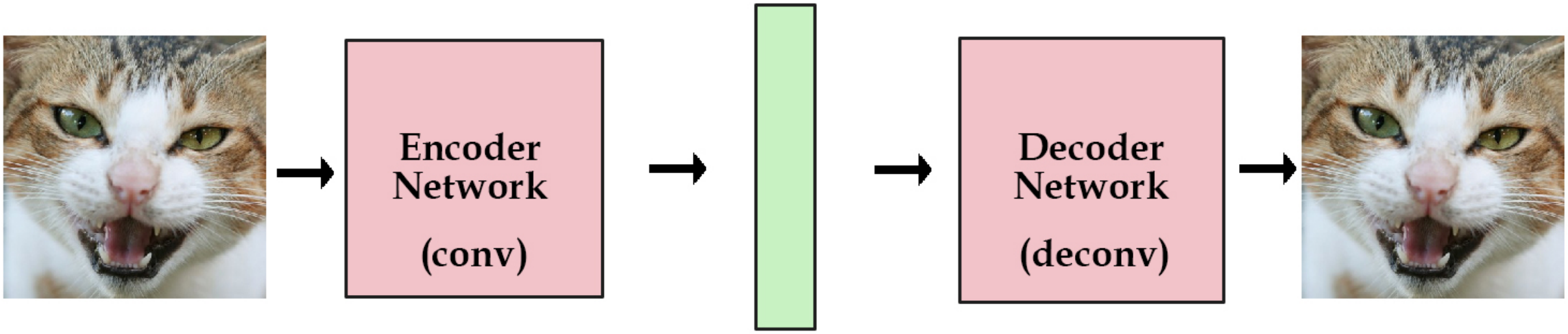
Autoencoder

Can we extract the feature without given labels?



Autoencoder

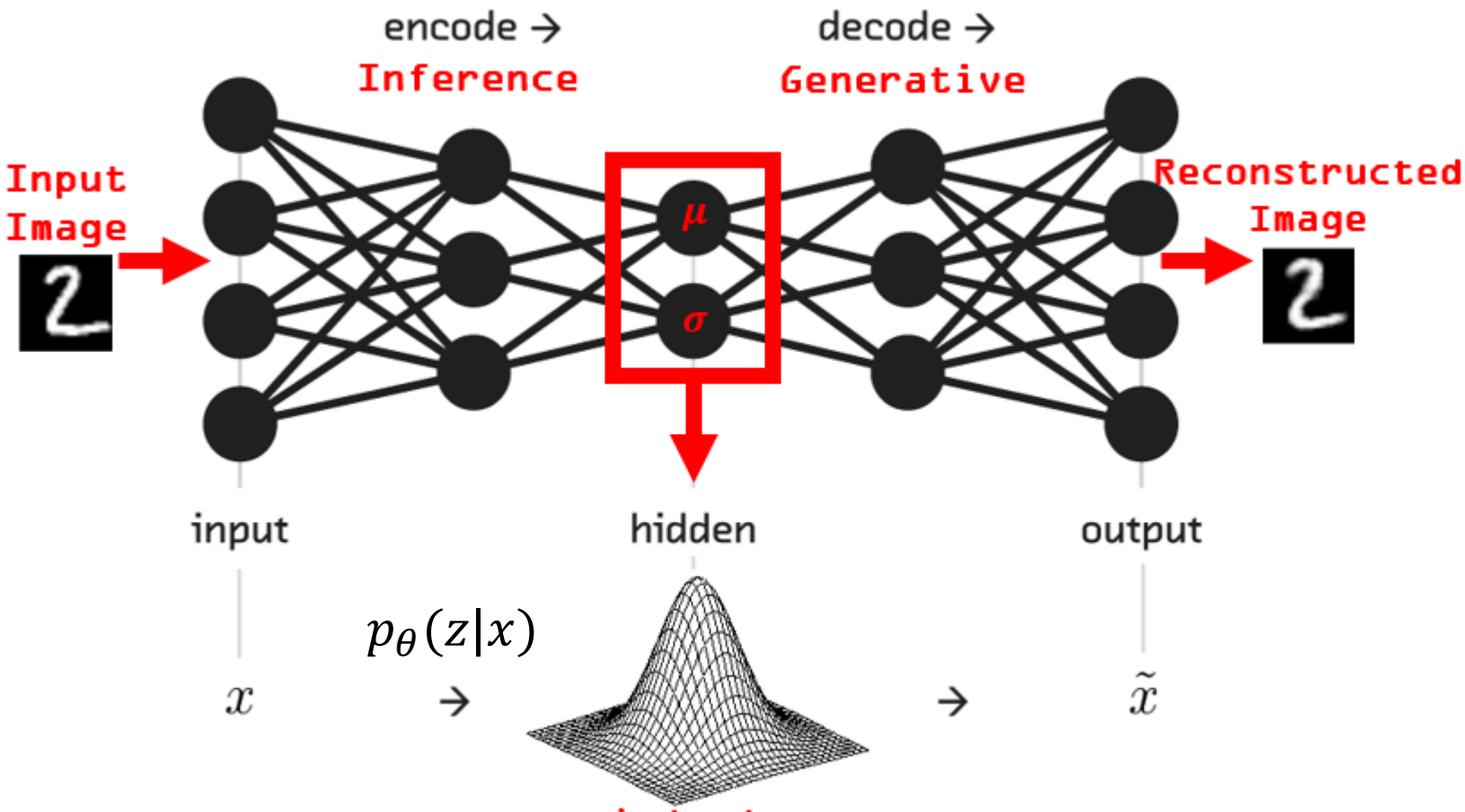
Can we extract the feature without given labels?



latent vector / variables

: contains the essential information of given input

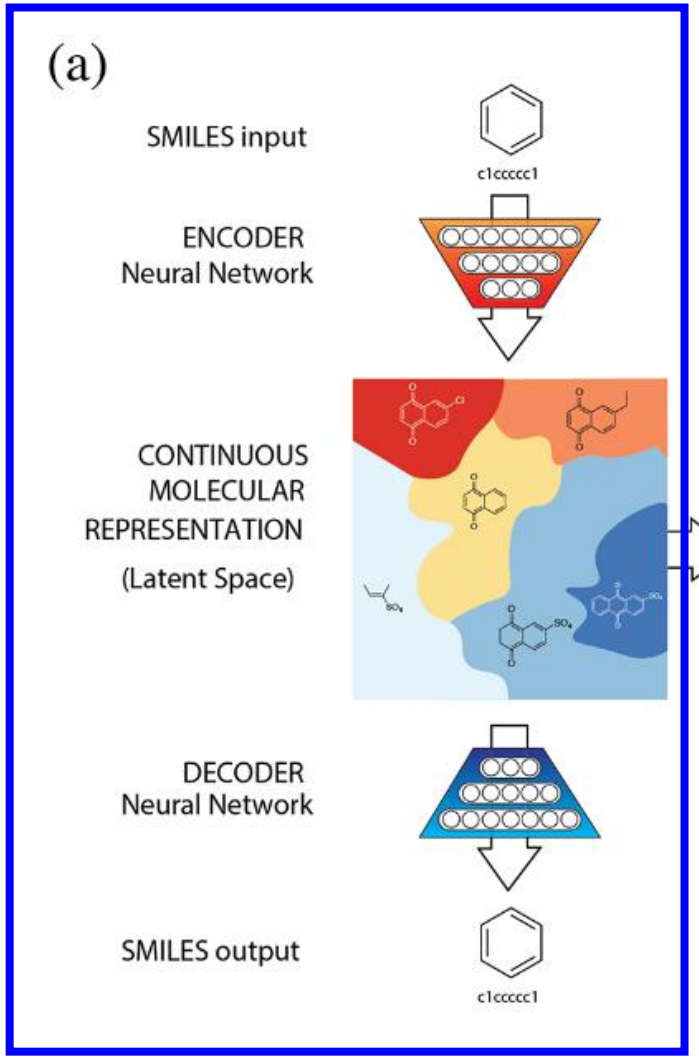
Variational autoencoder

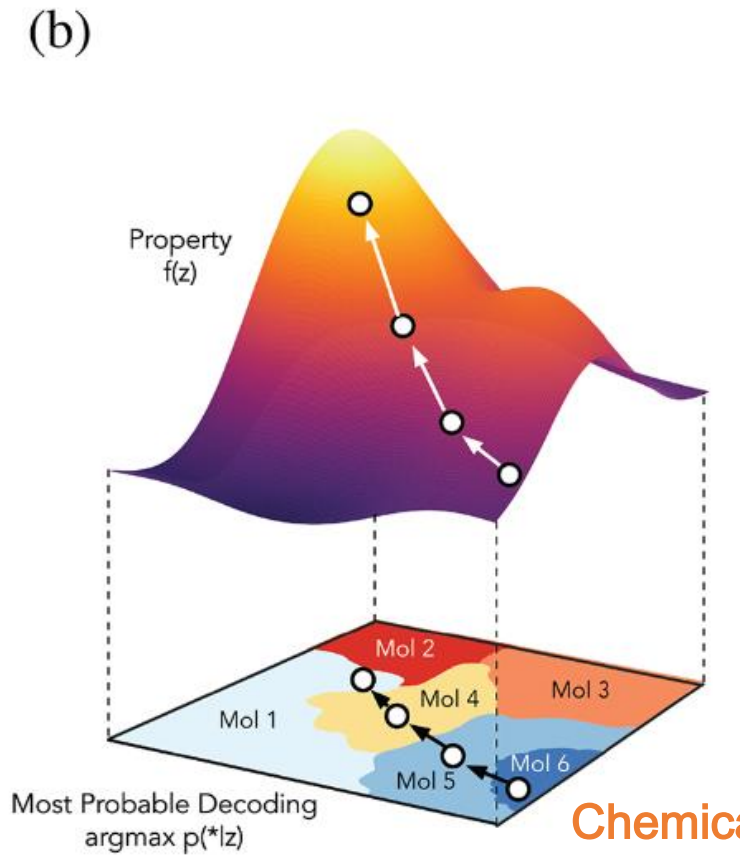


1) Latent vector has a distribution
2) This distribution is approximated with $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$

Variational autoencoder

ChemicalVAE



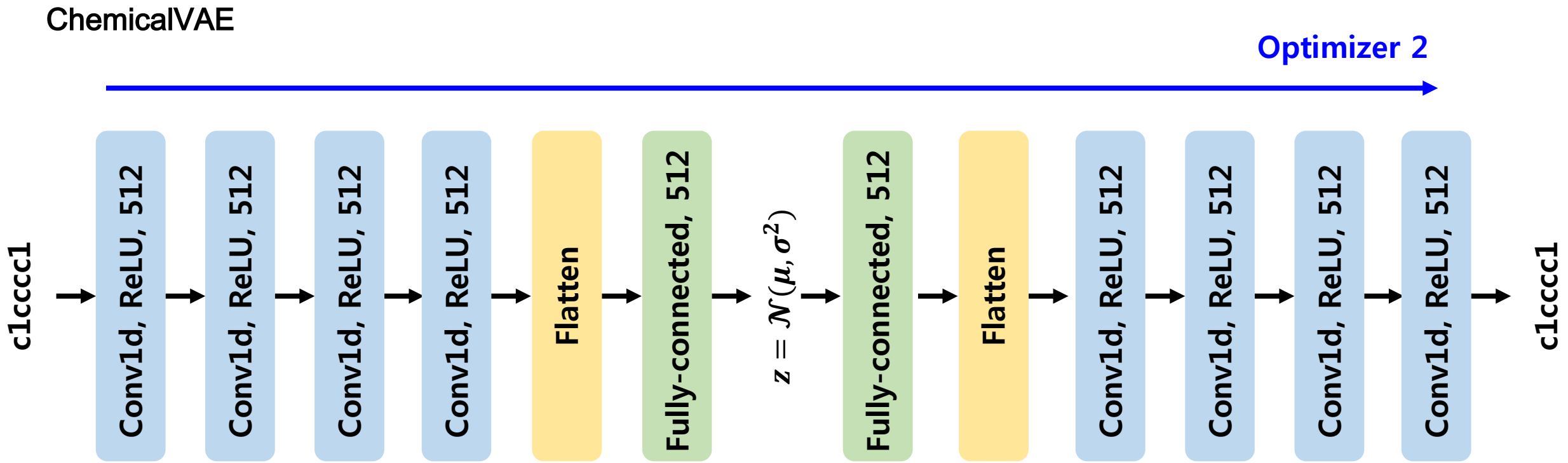


Chemical space generated by VAE

Interest of last practice

Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *ACS central science* 4.2 (2018): 268-276.

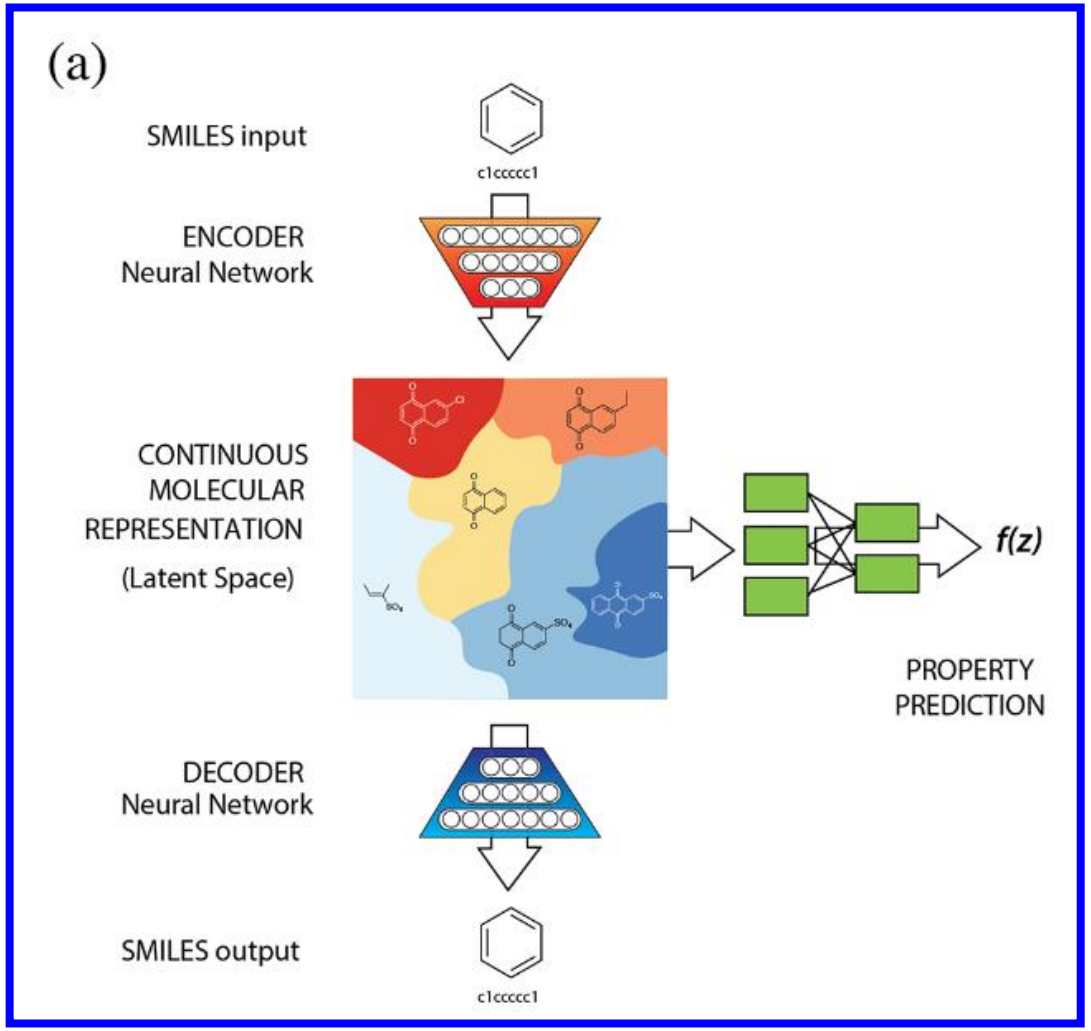
Variational autoencoder

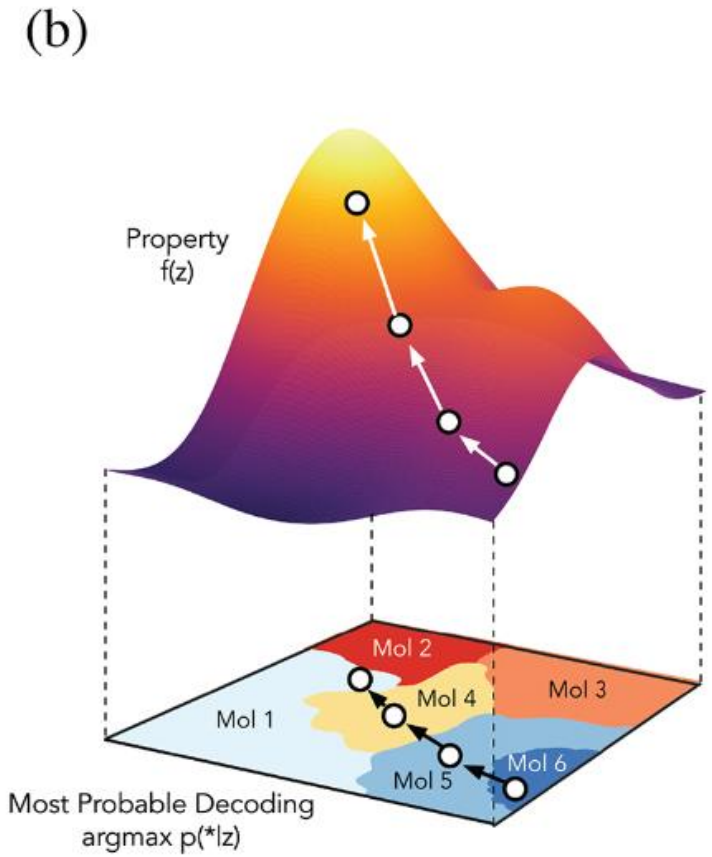


$$\mathcal{L} = \mathcal{L}_{reconstr} = \sum_i \|x_i - \hat{x}_i\|^2$$

Variational autoencoder

ChemicalVAE

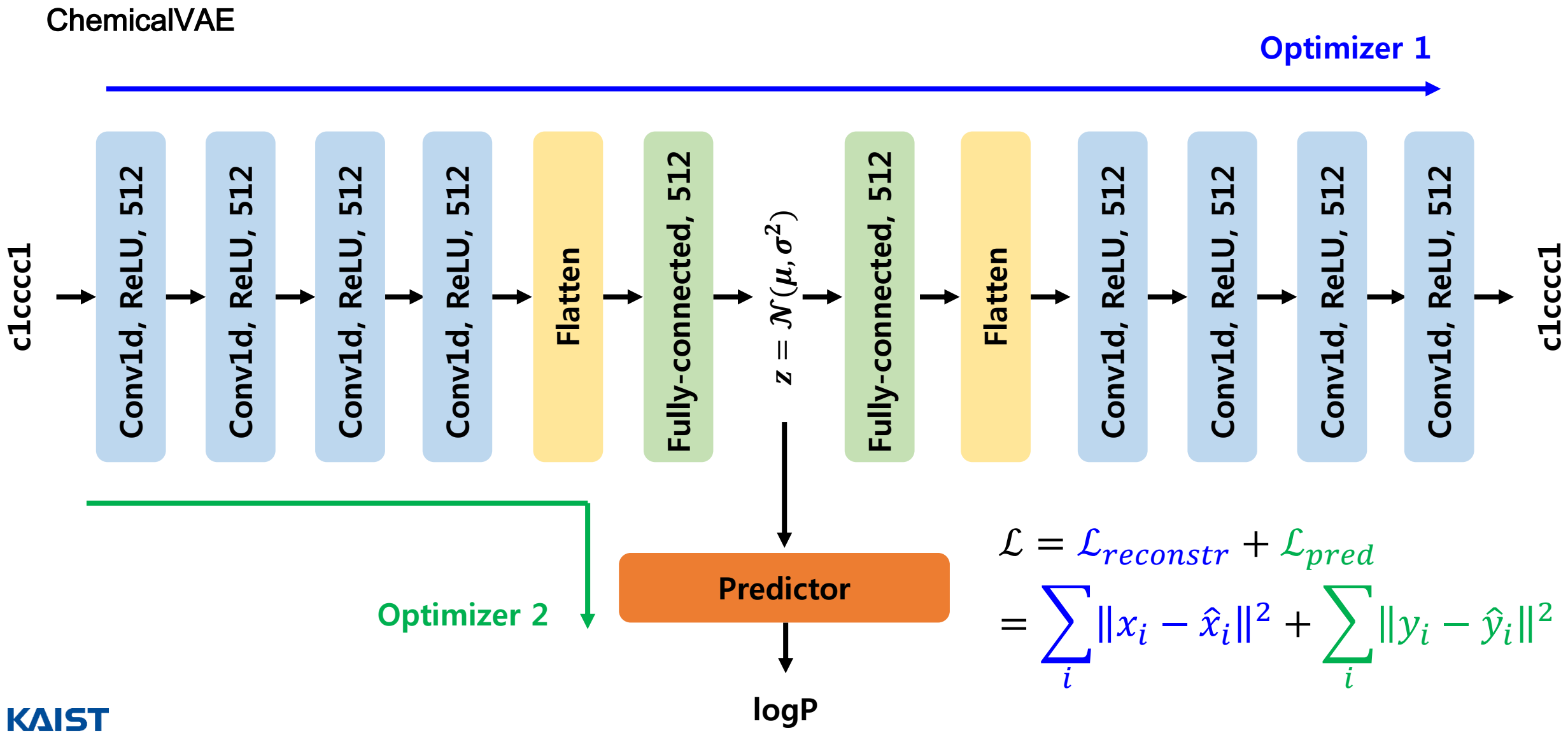




Interest of this practice

Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *ACS central science* 4.2 (2018): 268-276.

Variational autoencoder

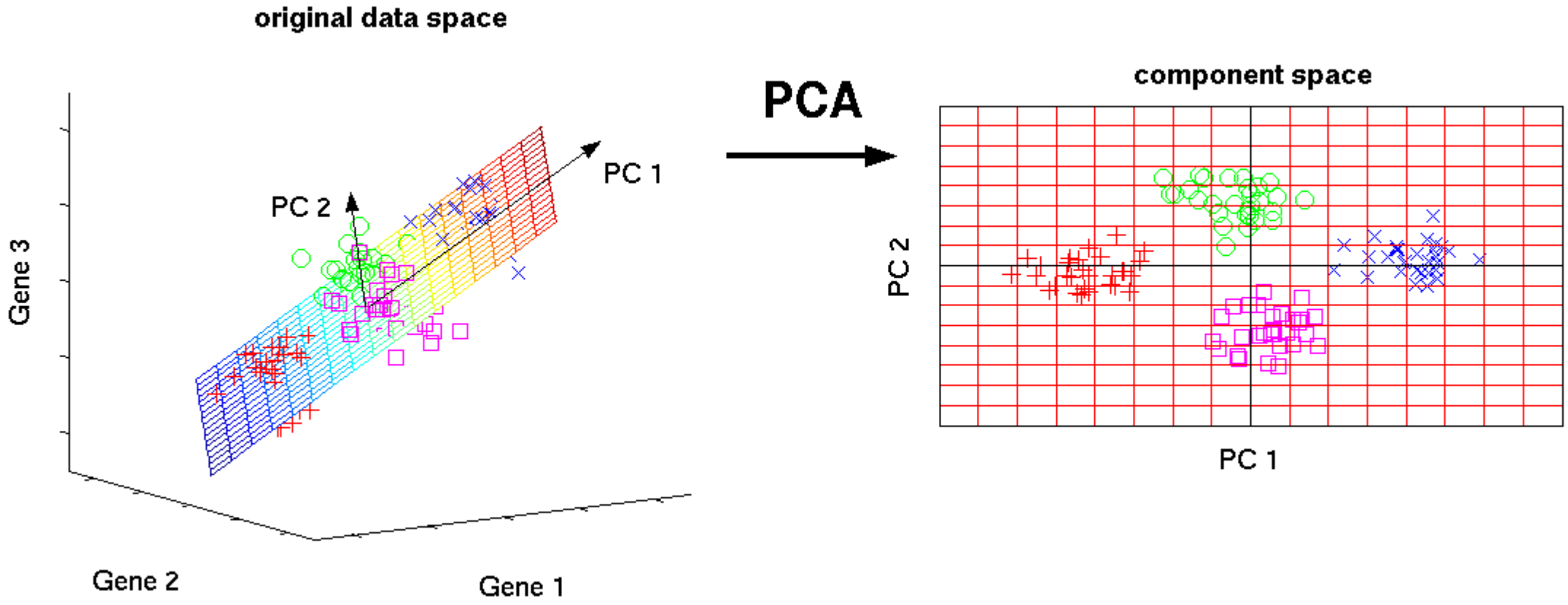


$$\mathcal{L} = \mathcal{L}_{reconstr} + \mathcal{L}_{pred}$$

$$= \sum_i \|x_i - \hat{x}_i\|^2 + \sum_i \|y_i - \hat{y}_i\|^2$$

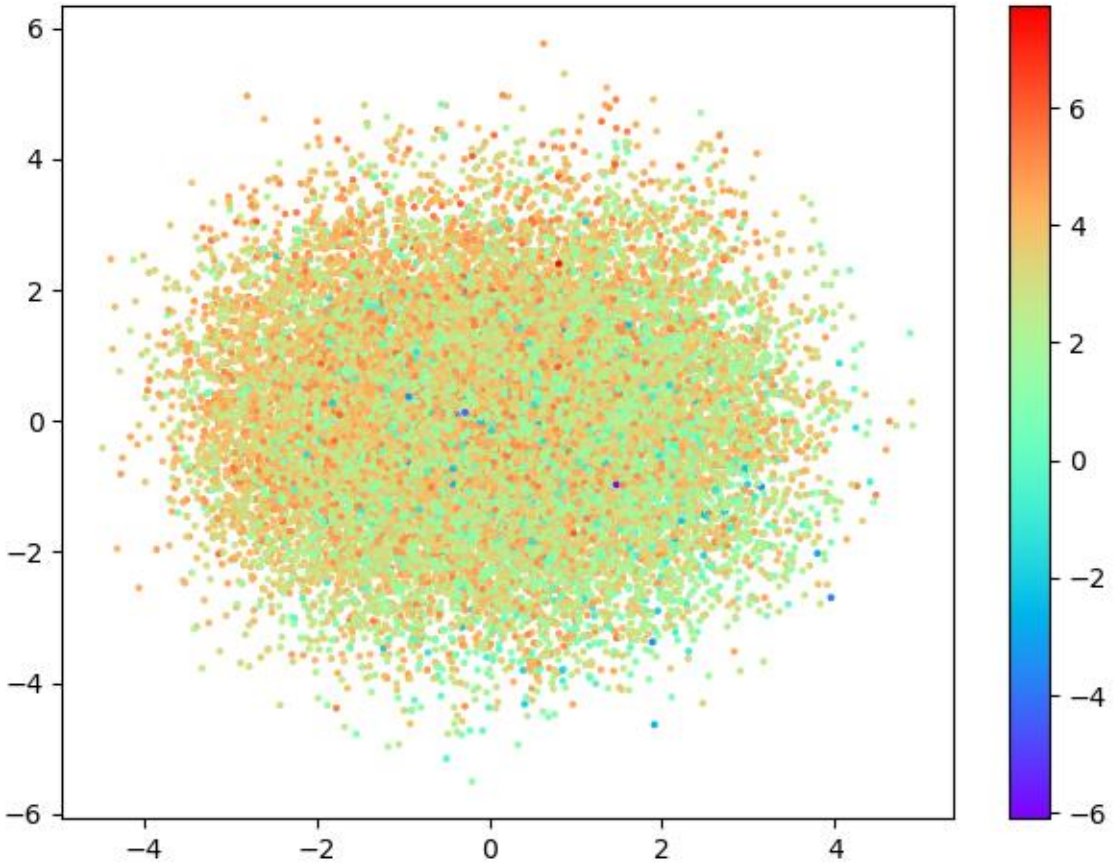
KAIST

Principal Component Analysis

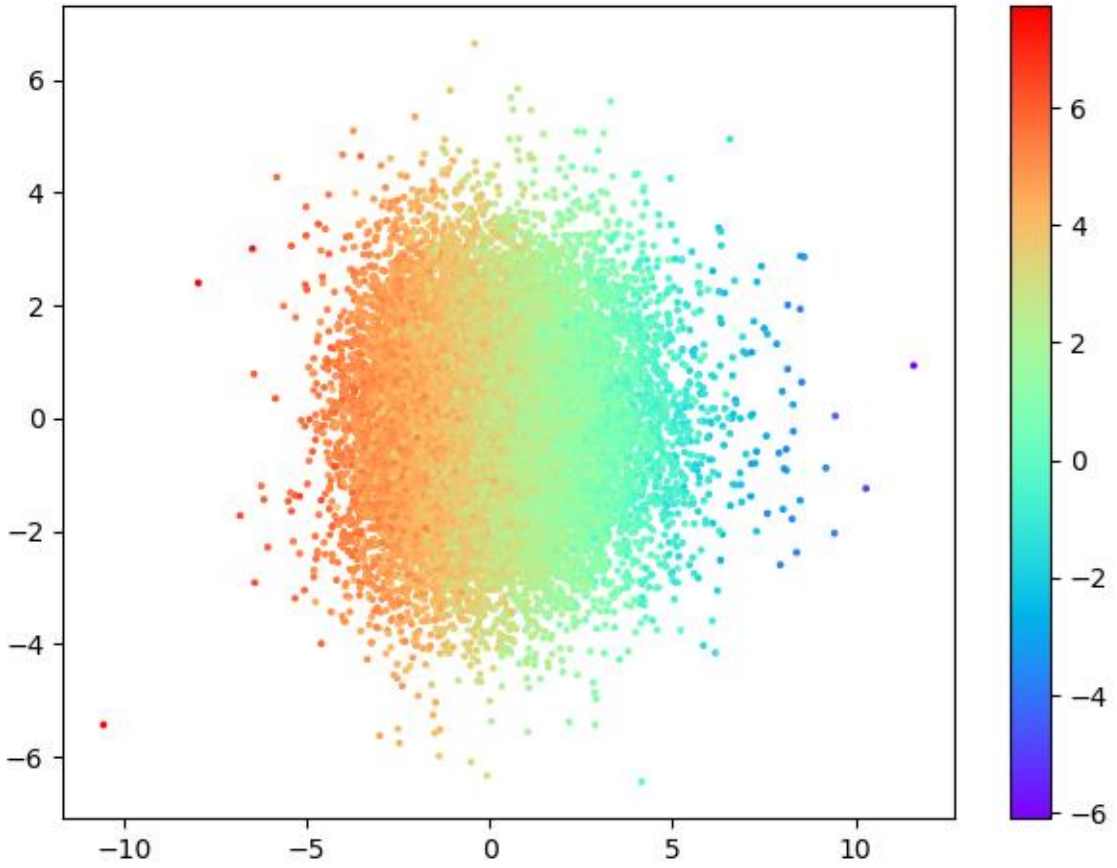


Principal Component Analysis

VAE



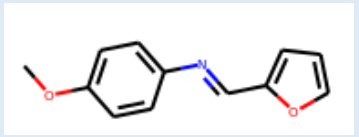
VAE + jointly training

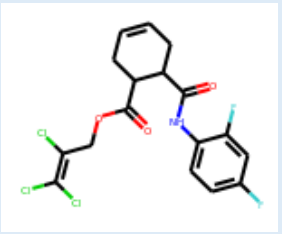


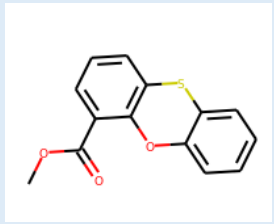
Jointly training makes latent vectors of molecules having similar property closer in the latent space.

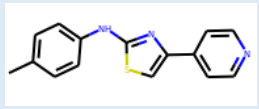
KAIST

Latent space searching







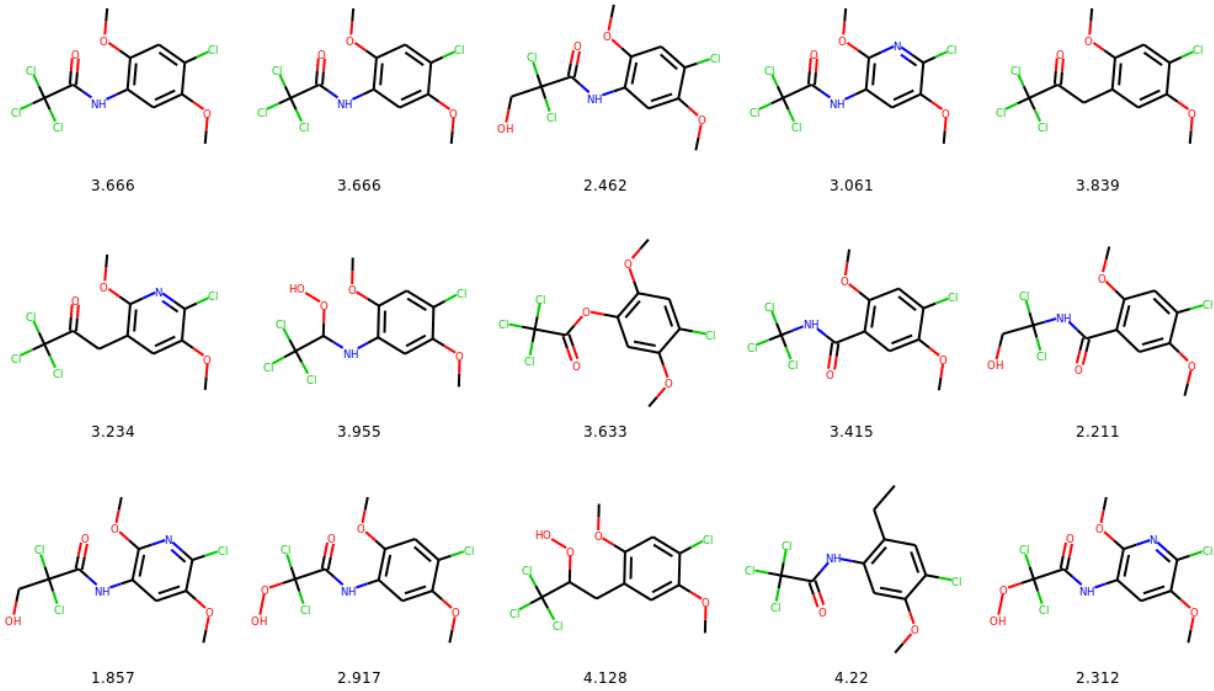


Latent space

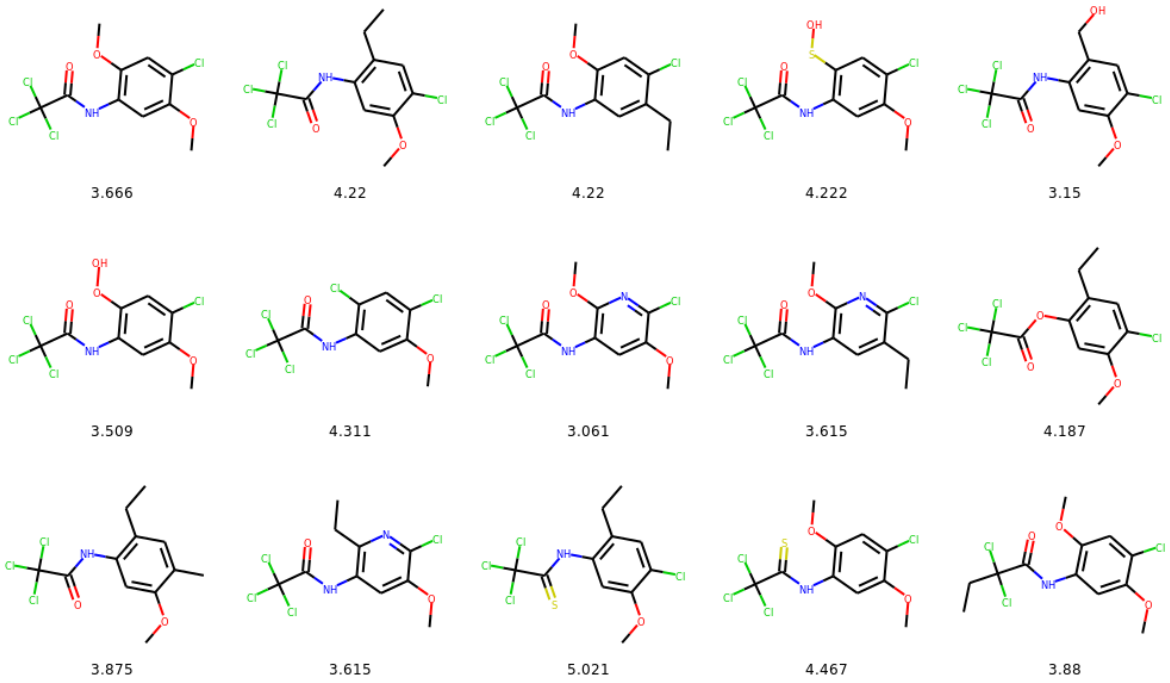
$\sigma = 0.1$

Latent space searching

VAE



VAE + jointly training

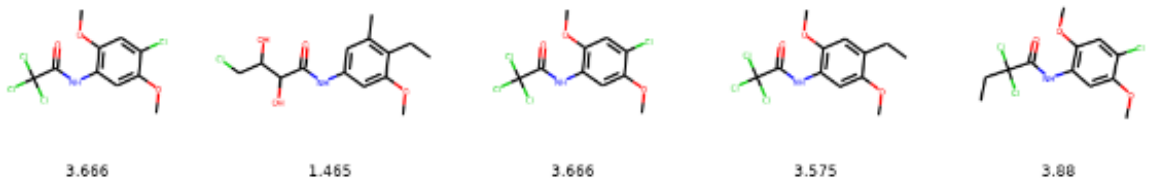


KAIST

$\sigma = 0.3$

Latent space searching

VAE



VAE + jointly training