Bayesian optimization

Hyperparameter tuning

Network parameters:

- Number of layers
- Layer sizes
- Dropout: on/off
- Batch normalization: on/off
- Nonlinearity

Training parameters:

- Learning rate
- Momentum

Usually finds better optima than when tuned by hand

Honest comparison with other methods in research

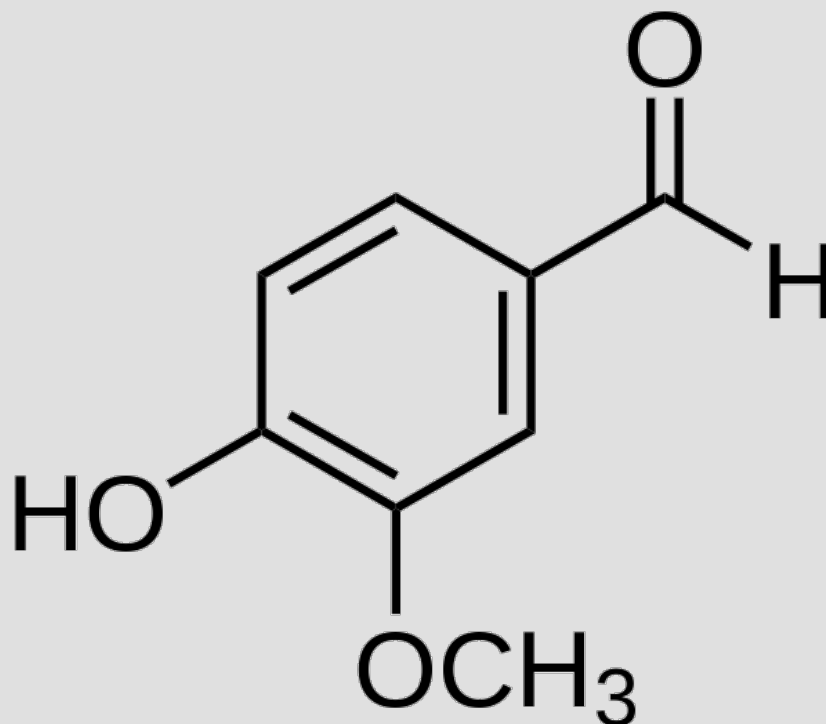# Discrete and continuous variables

- Treat discrete variables as continuous when fitting process

- Maximize $\mu(x)$ for each possible value of discrete variables

- Multi-armed bandit: all variables are discrete

Drug discovery

string representation of a molecule
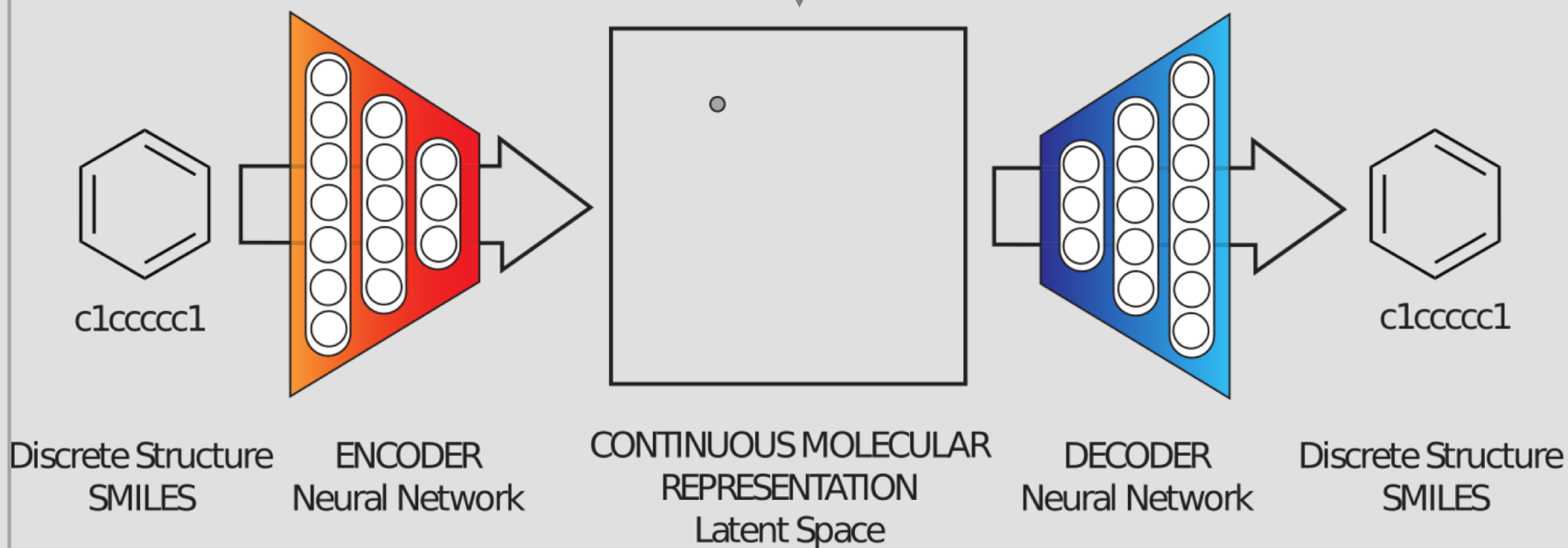
SMILES: COC1=C(C=CC(=C1)C=O)O

SMILES = Simplified Molecular Input Line Entry Specification
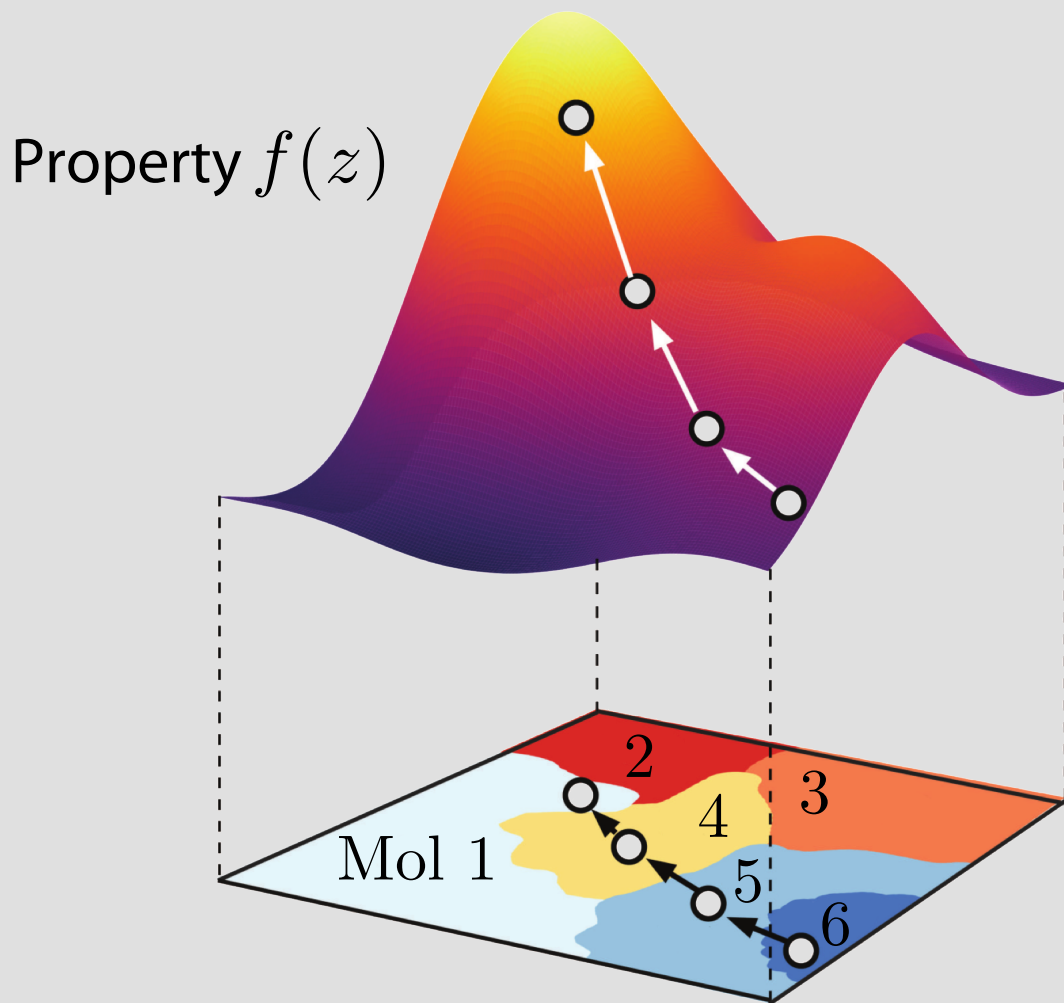
Encoding-decoding

Train VAE on SMILES:

Any point reconstructs valid molecule

c1ccccc1

Discrete Structure
SMILES

ENCODER
Neural Network

CONTINUOUS MOLECULAR
REPRESENTATION
Latent Space

DECODER
Neural Network

c1ccccc1

Discrete Structure
SMILES

Rafael Gómez-Bombarelli et. al. https://arxiv.org/abs/1610.02415

Bayesian optimization

Property $f(z)$

Optimize property like effectiveness against cancer.

While **True**:

1. Find maximum of acquisition function

2. Perform trials on new molecule

Rafael Gómez-Bombarelli et. al. https://arxiv.org/abs/1610.02415