Deep methods for interpretable analysis of gene interactions

Clayton W. Seitz

February 23, 2022

Outline

The logic of generative modeling

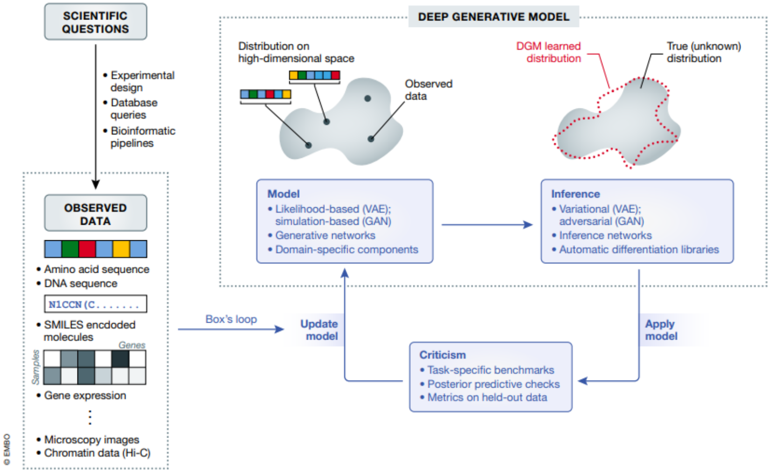
Say we have a set of variables $\mathbf{x} = (x_1, x_2, ..., x_n)$ which might have some statistical dependence

The variable $\mathbf{x}$ might be an amino acid sequence, gene expression data, microscopy image, etc.

- Often we are handed a batch of empirical samples $\{\mathbf{x}_i\}_{i=1}^{N}$
- We want to know the generating distribution $p(\mathbf{x})$

In supervised generative learning, we try to explicity learn the joint distribution $p(\mathbf{x}) = \prod_{i=1}^{N-1} p(x_i|x_{i+1:N})p(x_N)$, which is generally more difficult than discriminative learning.

Generative modeling with feedback



Perks of generative modeling

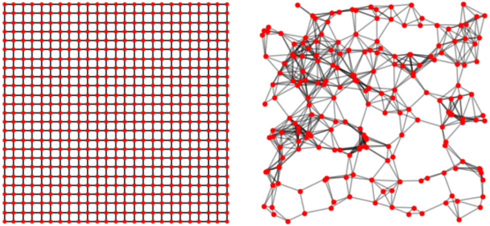
- Fitting complete multivariate distributions $p(\mathbf{x})$ goes beyond correlation-based or clustering approaches
- Correlations cannot discover partial correlation in the context of other neighbors
- Fitting $p(\mathbf{x})$ permits sampling based inference

Gene interactions are naturally represented as graphs

Quick review of convolutional neural networks

Assume stationarity of image statistics and locality of pixel dependencies

Defining Graph Convolution



Can be viewed as a generalization of convolutional neural networks to graph structured data

$$y = mx + b$$

Learning Bayesian network structure

Directed graph convolution on Bayesian networks

References I