A (very) brief introduction to graphical models

Clayton W. Seitz

February 16, 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.

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 inference

Why generative modeling is difficult

When describing a distribution over multiple variables, we may not know the proper normalization Z. That is,

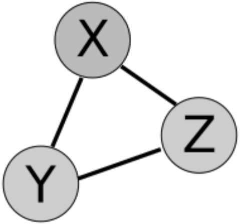
$$p(\mathbf{x}) = \frac{1}{Z}\tilde{p}(\mathbf{x})$$

This very important situation arises in several contexts:

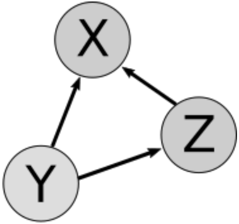
1. In Bayesian inference where $p(x_1|x_2) = p(x_2|x_1)p(x_1)/p(x_2)$ is intractable due to $Z = p(x_2) = \int p(x_2|x_1)p(x_1)dx_1$. This integral can be very difficult or impossible to compute.

2. In models from statistical physics, e.g. the Ising model, we only know $\tilde{p}(\mathbf{x}) = e^{-H(\mathbf{x})}$ where $H(\mathbf{x})$ is the Hamiltonian

Primary types of graphical models



Markov Random Field

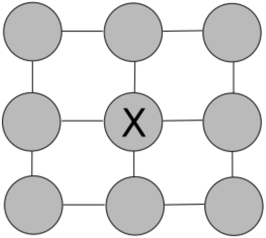


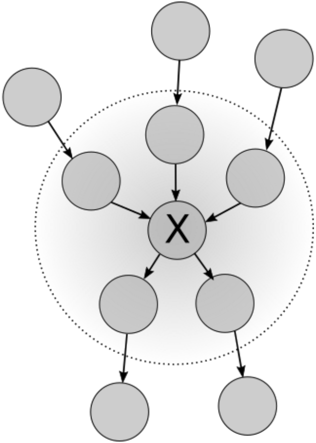
Bayesian Network

MRF: $P(X, Y, Z) = \psi(X, Y)\psi(X, Z)\psi(Y, Z)$

Bayes: $P(X, Y, Z) = P(X|Y, Z)P(Z|Y)P(Y)$

The Markov Blanket





Bayesian networks for modeling gene interactions

MCMC Structure Samplers

Bayesian image reconstruction

Say a fluorophore emits photons at a rate λ_n. This is the best we can do according to QM

For a CMOS array with quantum efficiency $\gamma\ [e^-/p]$ we have

$$I_n = \gamma g_n P_n(\lambda_n) + G_n(\mu_n; \sigma_n^2) + \beta$$

where $\mu_n\ [\mathrm{ADU}]$ is the detector offset and $g_n\ [\mathrm{ADU}/e^-]$ is the gain.

All we know is λ_n, so both the true signal I_n and the detected signal $\hat{I}_n$ are stochastic processes.

$$P_\lambda(I_n, \hat{I}_n) = \frac{1}{Z}\frac{\exp\left(-\lambda_n\right)\lambda_n^p}{p!}\exp\left(-\frac{(D - g_n p - \mu_n)^2}{\sigma_n^2}\right)$$

Bayesian image reconstruction

Marginalizing over p gives the noise model as a function of the rate λ_n

$$P_\lambda(I_n) = \frac{1}{Z}\sum_p \frac{\exp\left(-\lambda_n\right)\lambda_n^p}{p!}\exp\left(-\frac{(D - g_n p - \mu_n)^2}{\sigma_n^2}\right)$$

References I