Autoregression Model

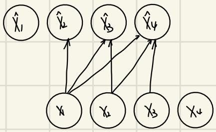
1. pick an ordering of all random variables eg $(X_1, X_2 \cdots X_{784})$
2. without loss of generality, use chain rule

$$p(X) = p(X_1) \cdot p(X_2|X_1) \cdots p(X_{784}|X_1 \cdots X_{783})$$

3. Assume

$$P(X_1 \cdots X_{784}) = P_{cpt}(X_1 ; \alpha_1) \cdot P_{logit}(X_2|X_1 ; \alpha_2) \cdots P_{logit}(X_n|X_1 \cdots X_{n-1} ; \alpha_n) \qquad P(X_2|X_1 ; \alpha_2) = \sigma(X_1 \cdot \alpha_2[1] + \alpha_2[0])$$

the logistic form is a modeling assumption



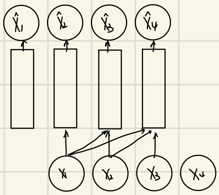
Fully Visible Sigmoid Belief Network

$$\hat{X}_i = p(X_i = 1 | x_1, x_2 \cdots x_{i-1})$$

To evaluate the probability $p(X_1 \cdots X_{784})$, just multiply all conditionals

To sample from $P(X_1 \cdots X_{784})$, sample each conditionals in order

neural autoregressive density estimation (NADE): Use nn to replace logistic functions



to model general discrete distribution: use a softmax, rather than sigmoid

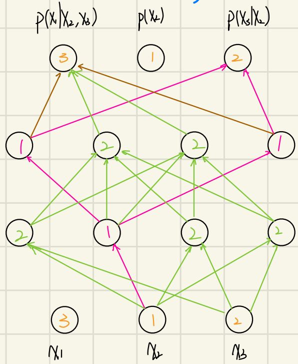
to model continuous distribution, eg. Gaussian $\quad p(X_i | X_1 \cdots X_{i-1}) \sim \mathcal{N}(u, \sigma)$, let $(u, \sigma) = \text{Neural Network}(X_1, \cdots X_{i-1})$

eg. mixture gaussian $\quad P(X_i | X_1 \cdots X_{i-1}) \sim \frac{1}{3}(\mathcal{N}(u_1, \sigma_1) + \mathcal{N}(u_2, \sigma_2) + \mathcal{N}(u_3, \sigma_3))$ let $(u_1, u_2, u_3, \sigma_1, \sigma_2, \sigma_3) = nn(X_1 \cdots X_{i-1})$

FVSBN and NADE looks like an autoencoder

A vanilla autoencoder is not a generative model, it doesn't define a distribution over X, and we cannot sample

· Masked Autoencoder for Density Estimation



pick an ordering for chain rule: $p(x_1, x_2, x_3) = p(x_2) \cdot p(x_3|x_2)\ p(x_1|x_2, x_3)$

$p(x_3|x_2)$ can only touch info from x_2

$p(x_1|x_2, x_3)$ can toch info from x_2 x_3

] autoregressive property

(mask some weights in an NN)

for each unit in a hidden layer, pick a random integer in [1, N-1]

for { hidden layers, / output layers, } connect to unit in the previous layer with { smaller or equal / strictly smaller } number