

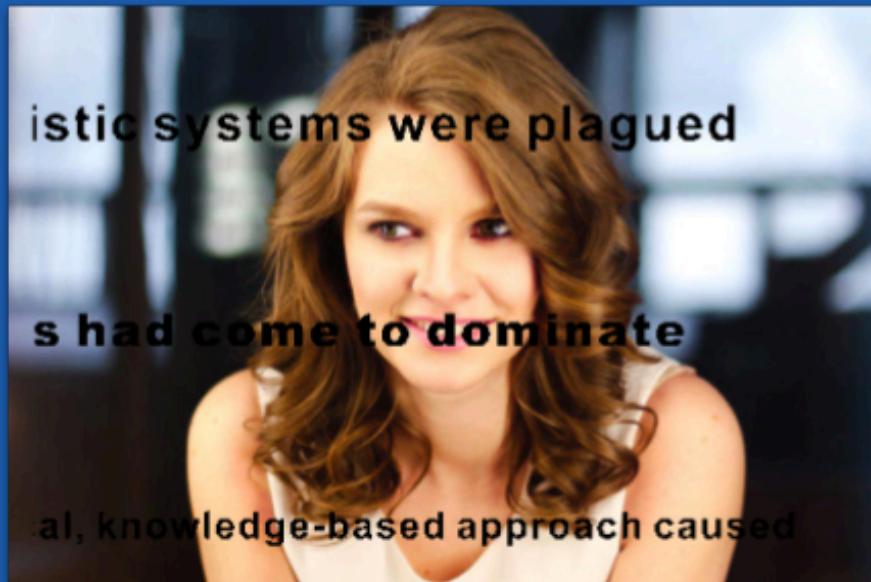
Deep Models for Distribution Estimation



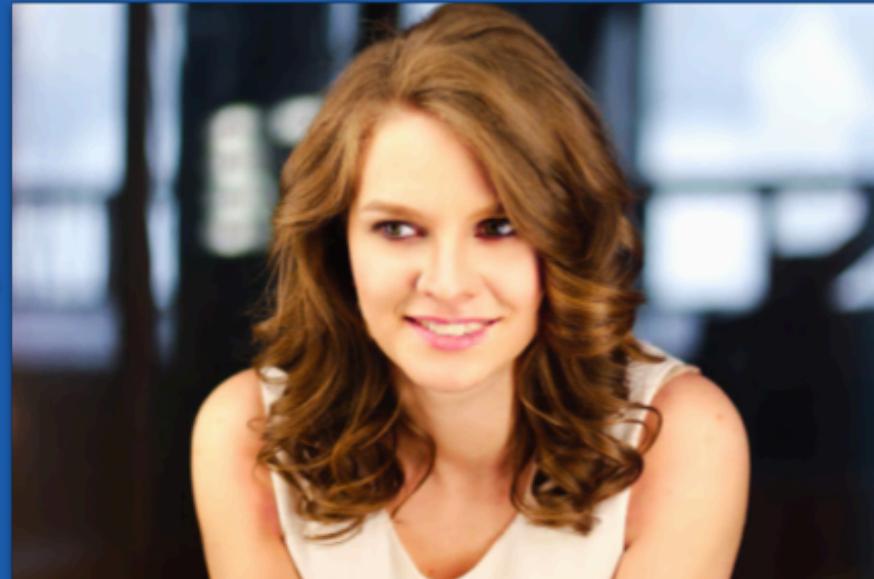
P

P

Inpainting



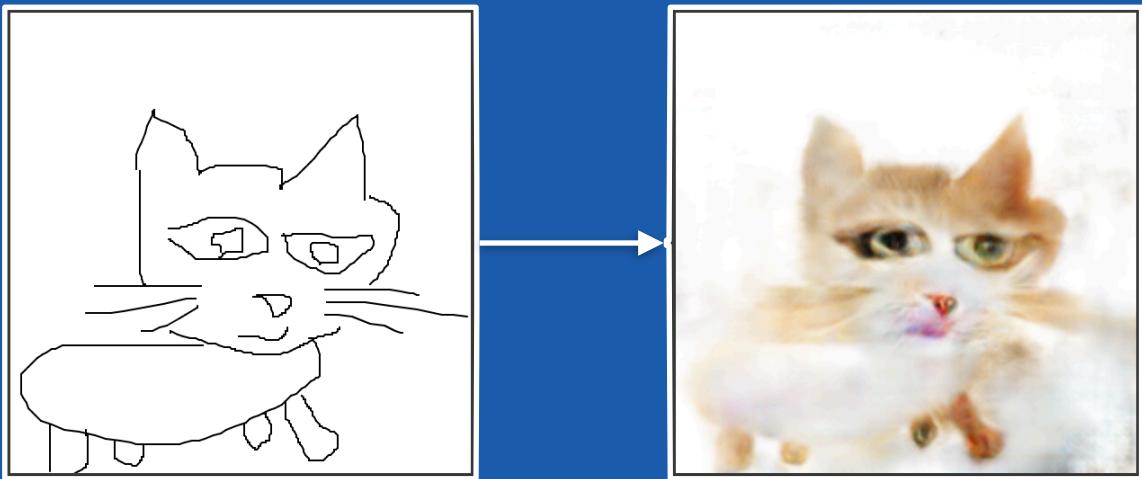
Corrupted



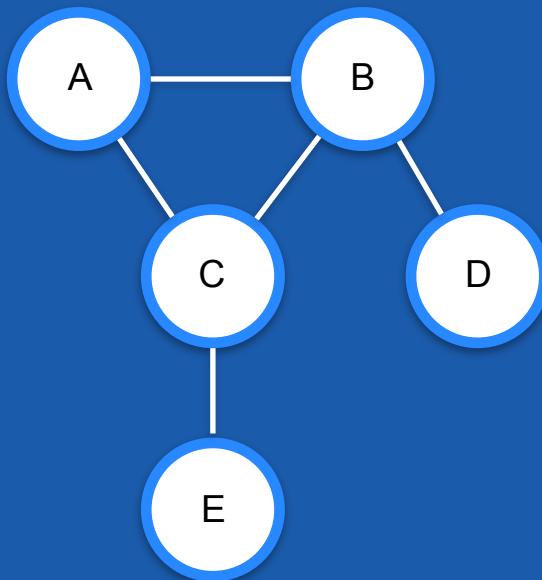
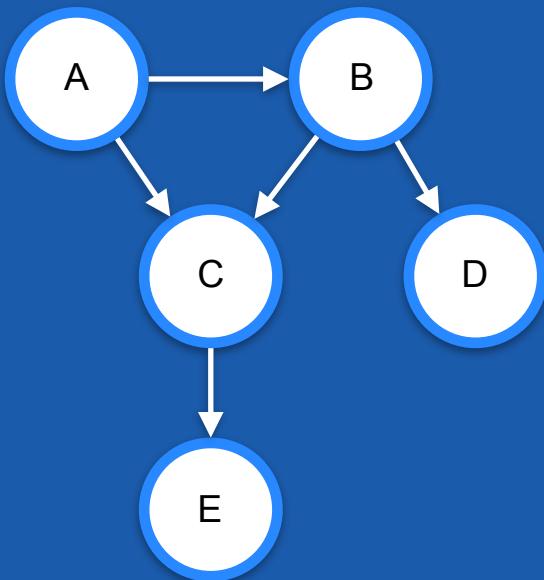
Deep Image Prior

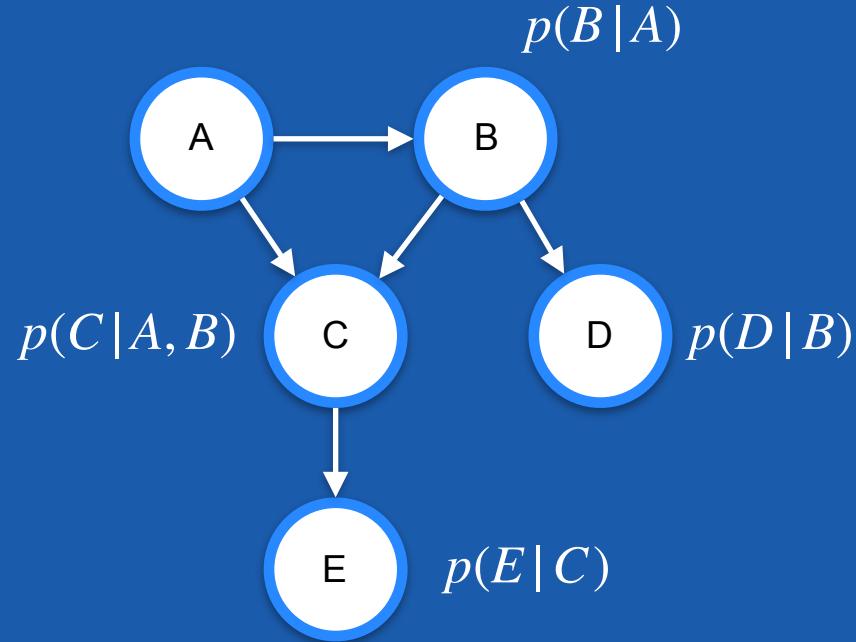
$$x \sim P$$



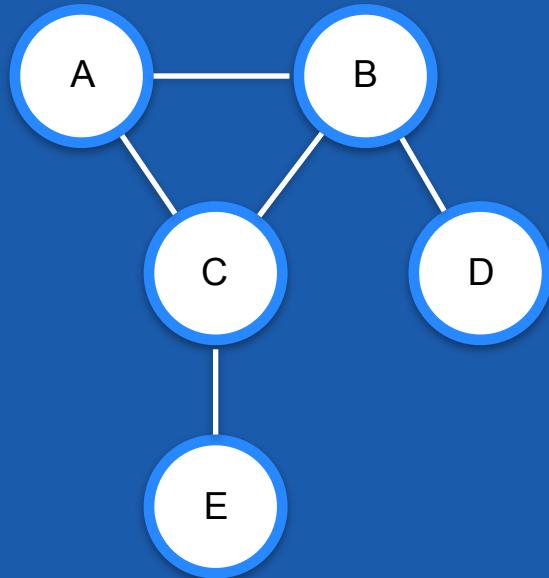
$$x \sim P$$


$p(x)$ $p($ $)$ 

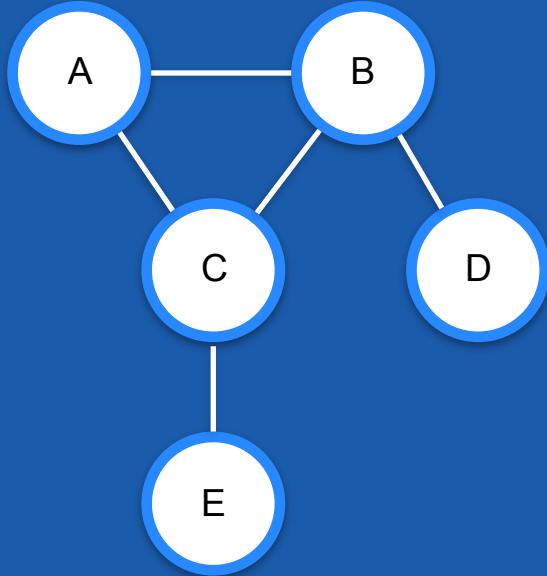




$$p(x) = \prod_i p(x_i | Pa(x_i))$$



$$\tilde{p}(x) = \prod_{C \in \mathcal{G}} \phi(C)$$



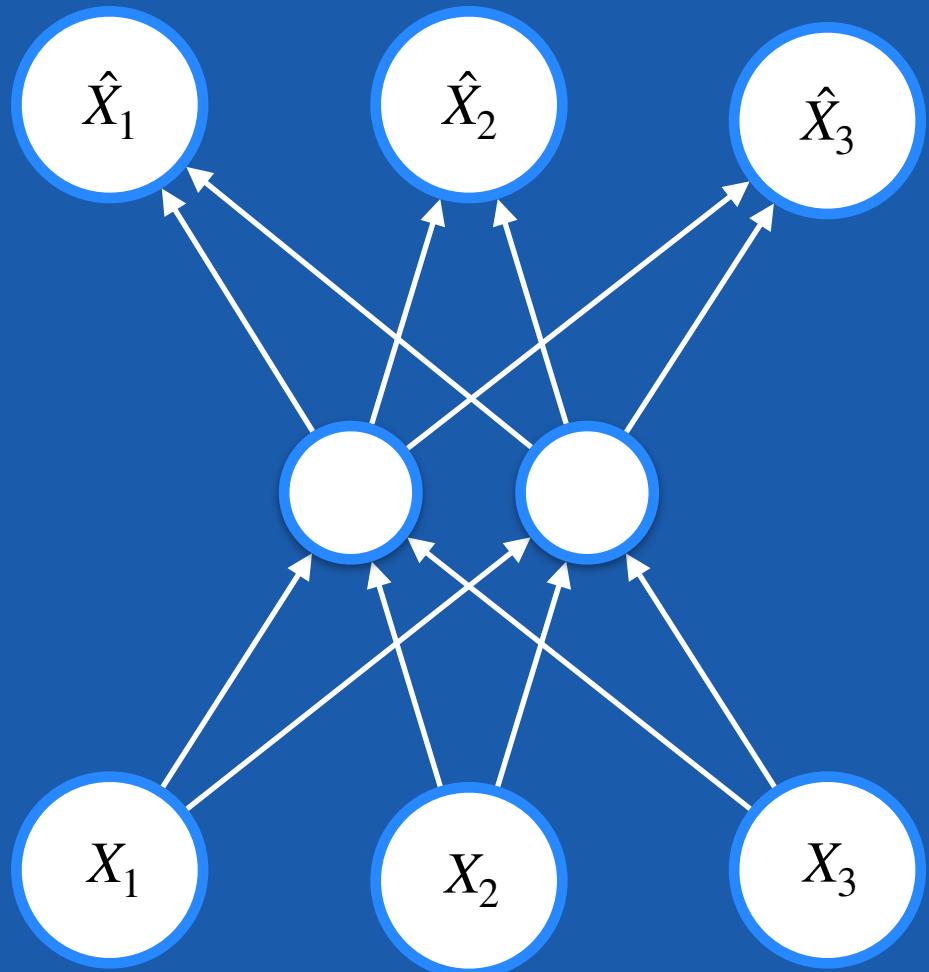
$$\tilde{p}(x) = \prod_{C \in \mathcal{G}} \phi(C)$$

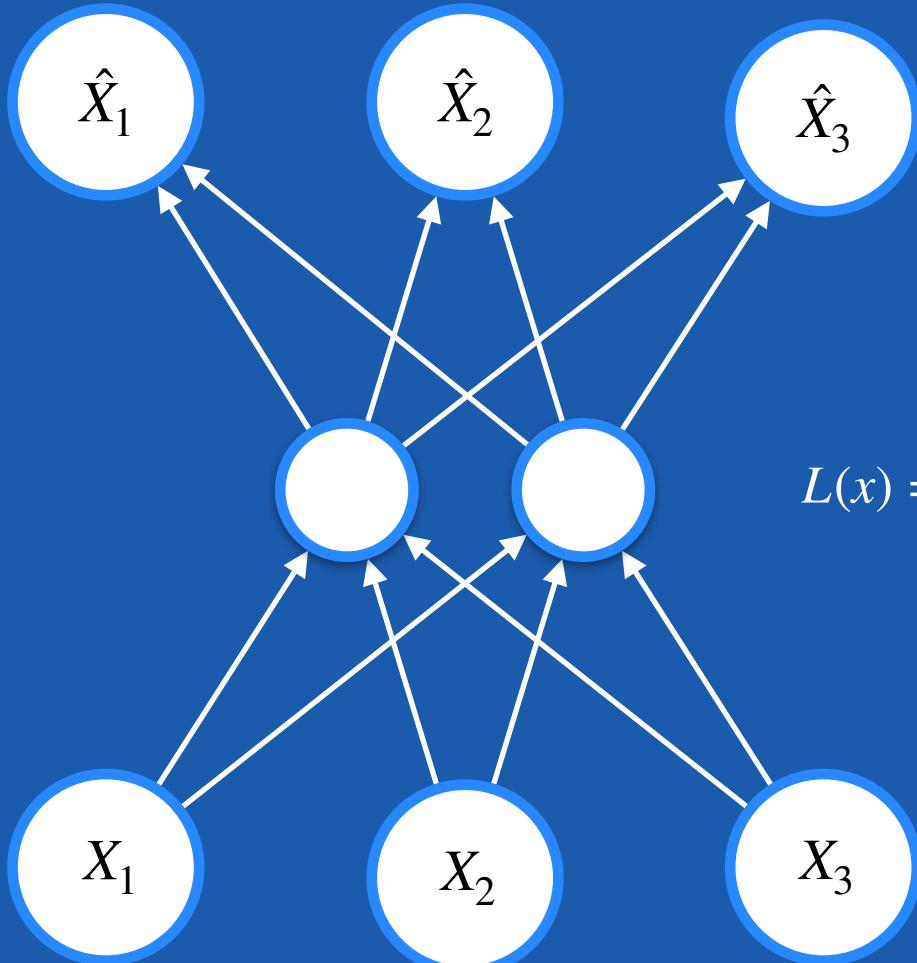
$$p(x) = \frac{1}{Z} \tilde{p}(x)$$

$$Z = \int \tilde{p}(x) dx$$

2784







$$\hat{X} = \text{sigm}(\dots)$$

$$L(x) = \sum_{d=1}^D -x_d \log \hat{x}_d - (1 - x_d)(1 - \log \hat{x}_d)$$

MADE: Masked Autoencoder for Distribution Estimation

Mathieu Germain

Université de Sherbrooke, Canada

MATHIEU.GERMAIN2@USHERBROOKE.CA

Karol Gregor

Google DeepMind

KAROL.GREGOR@GMAIL.COM

Iain Murray

University of Edinburgh, United Kingdom

I.MURRAY@ED.AC.UK

Hugo Larochelle

Université de Sherbrooke, Canada

HUGO.LAROCHELLE@USHERBROOKE.CA

$$p(x) = \prod_{d=1}^D p(x_d \mid x_{<d})$$

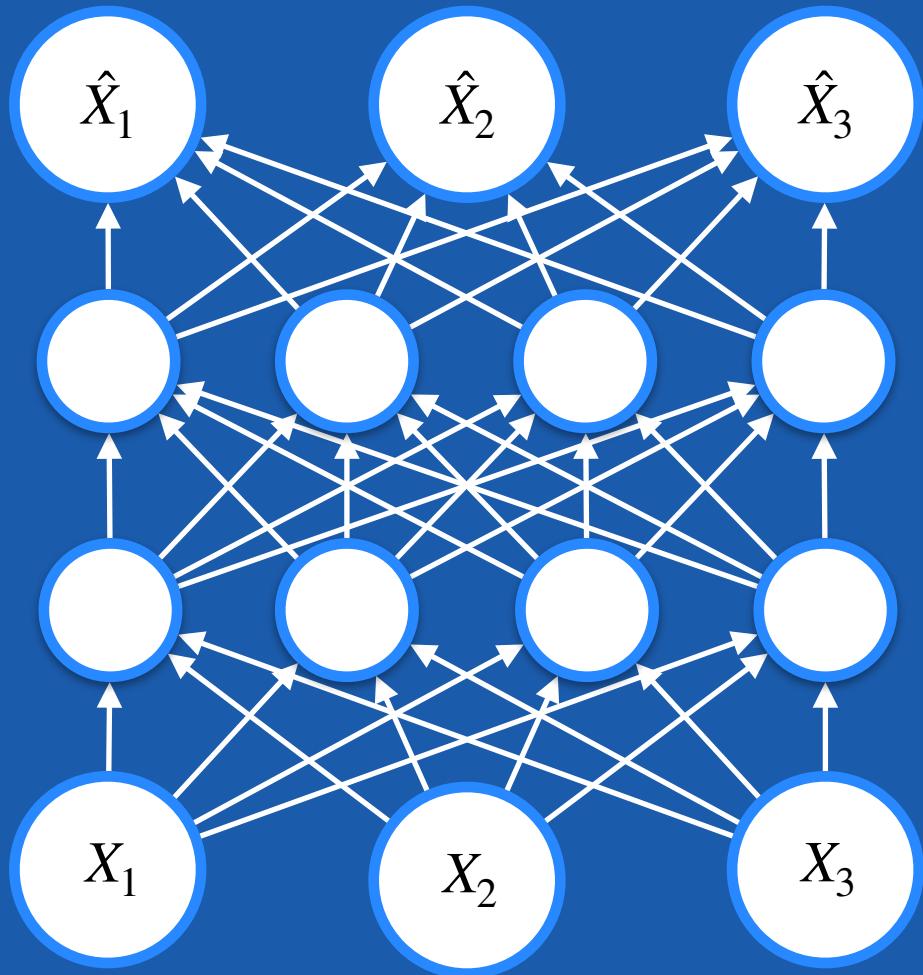
$$p(x) = \prod_{d=1}^D p(x_d | x_{<d})$$

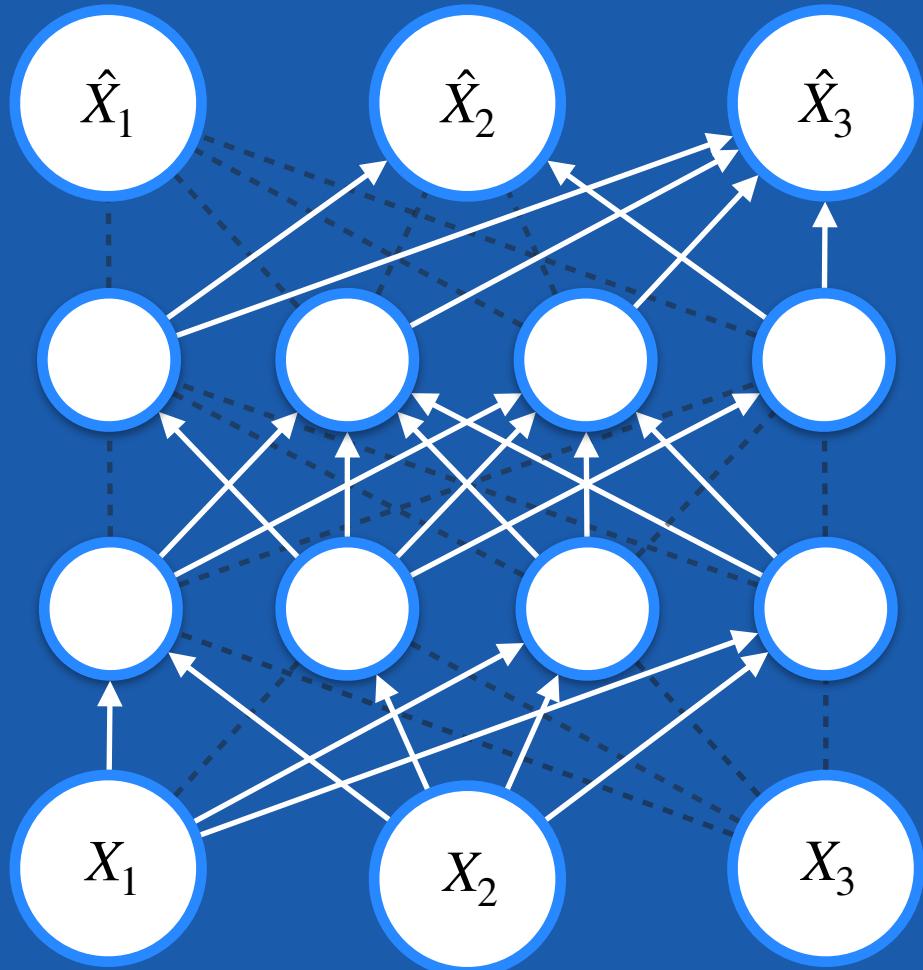
$$p(x_1) = p(x_1)$$

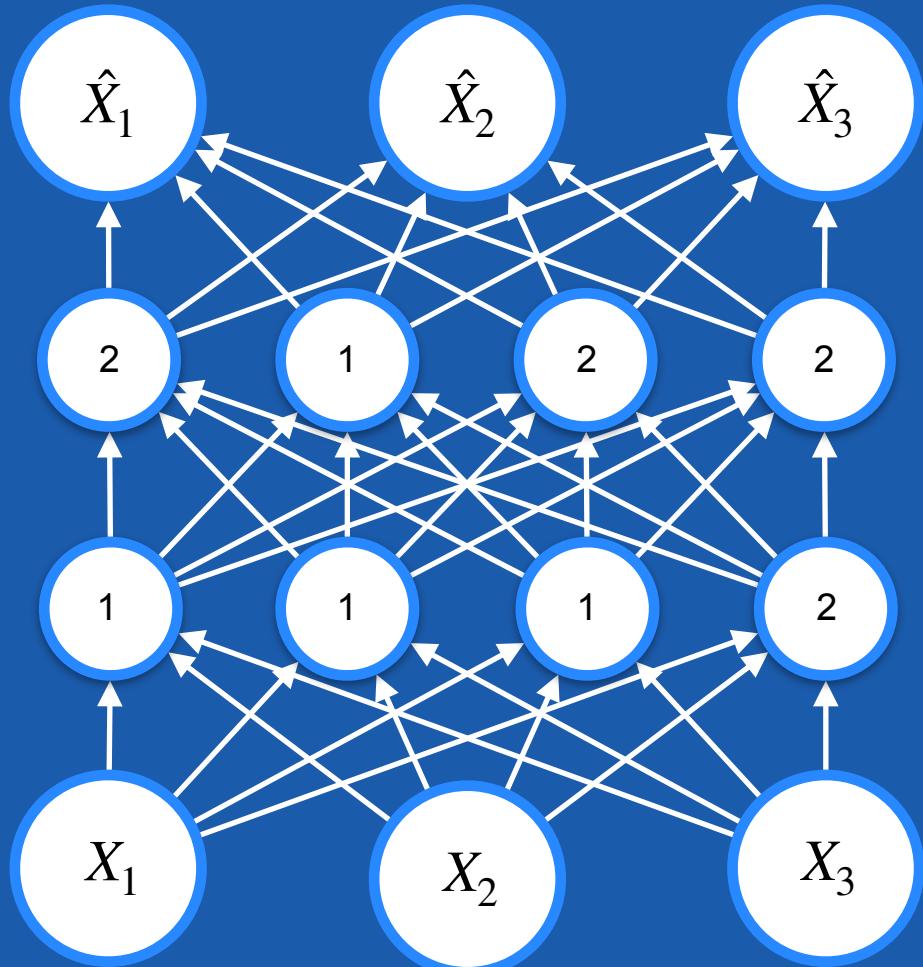
$$p(x_2) = p(x_2 | x_1)$$

$$p(x_3) = p(x_3 | x_1, x_2)$$

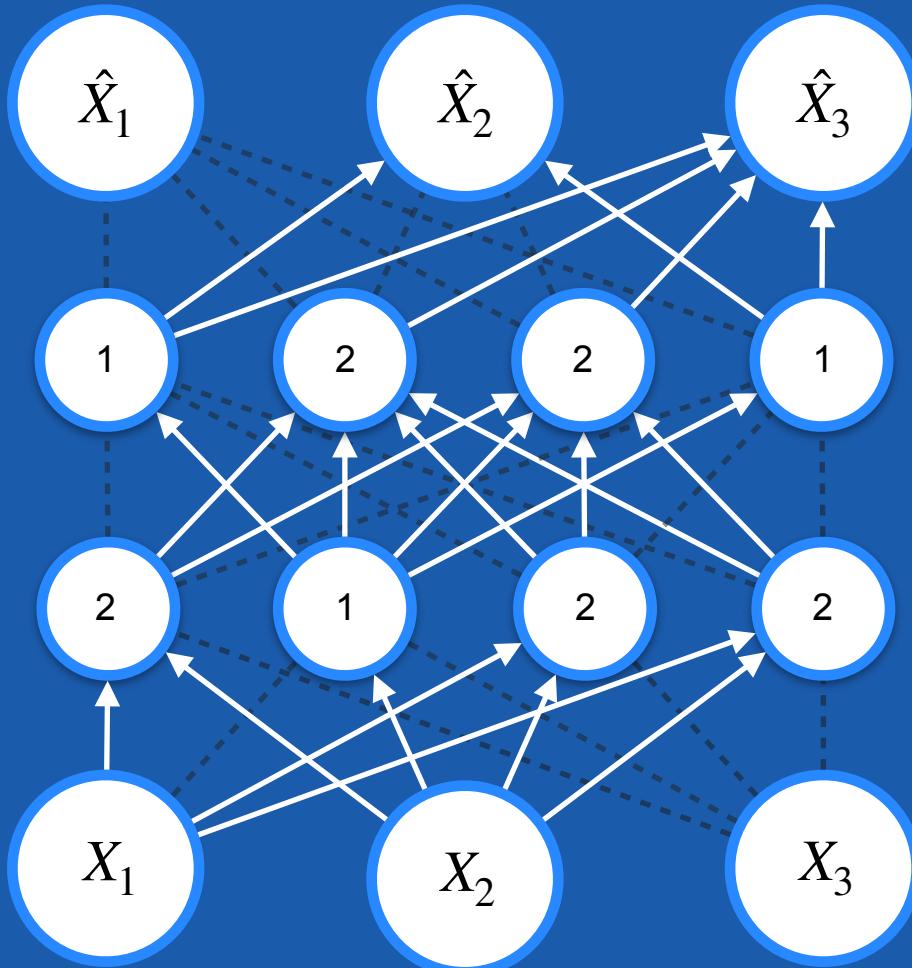
$$p(x) = p(x_3)p(x_2)p(x_1) = p(x_3 | x_1, x_2)p(x_2 | x_1)p(x_1)$$

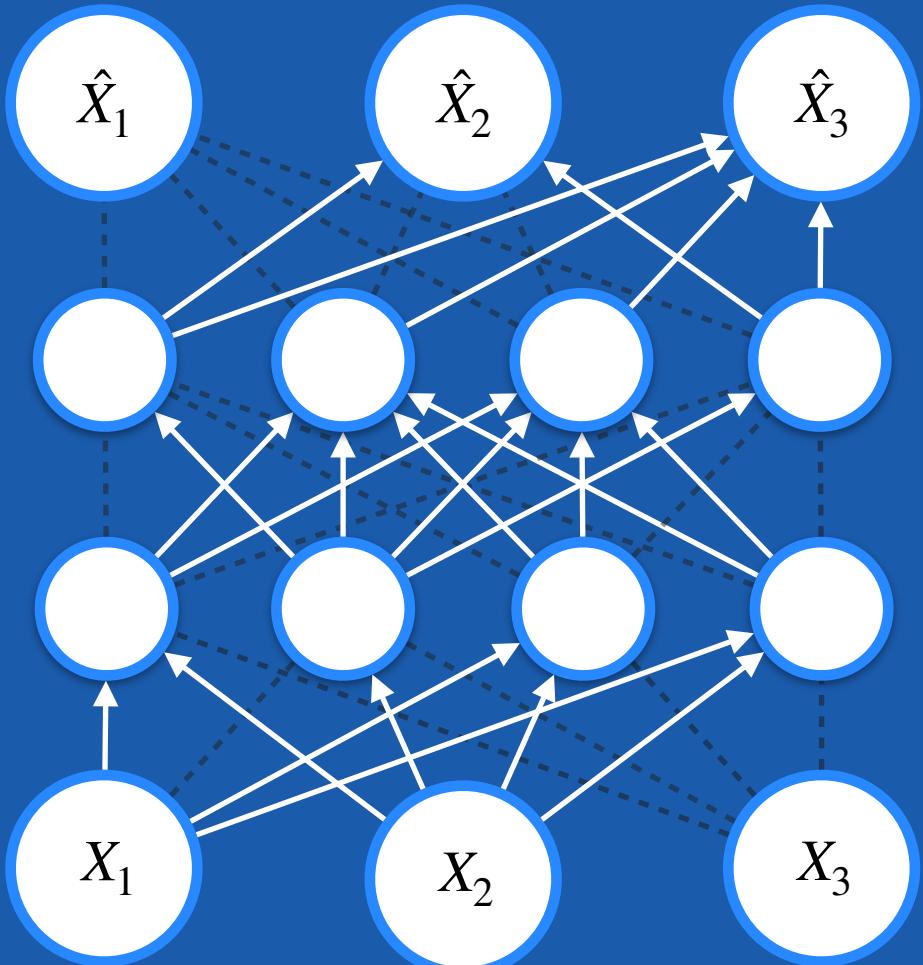






Node with order D can be connected only with nodes with same or smaller order





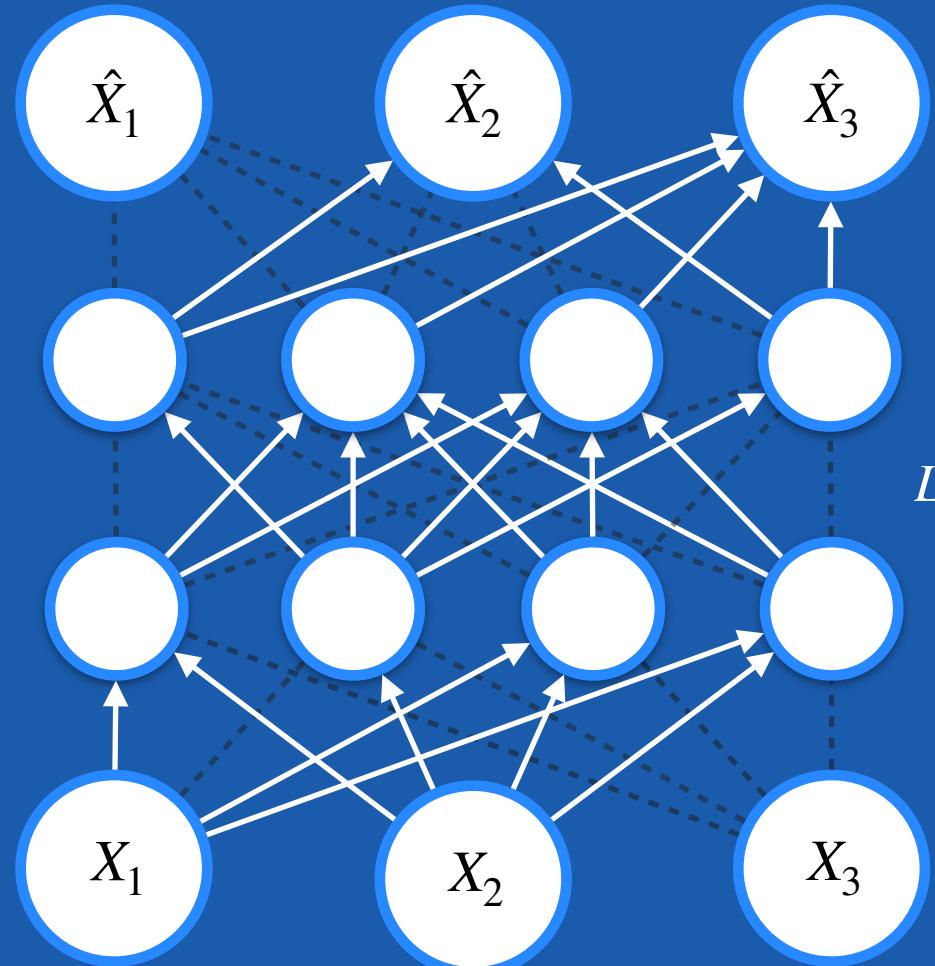
$$p(x_d = 1 | x_{<d}) = \hat{x}_d$$

$$p(x_d = 0 | x_{<d}) = 1 - \hat{x}_d$$

$$p(x) = \prod_{d=1}^D p(x_d | x_{<d})$$

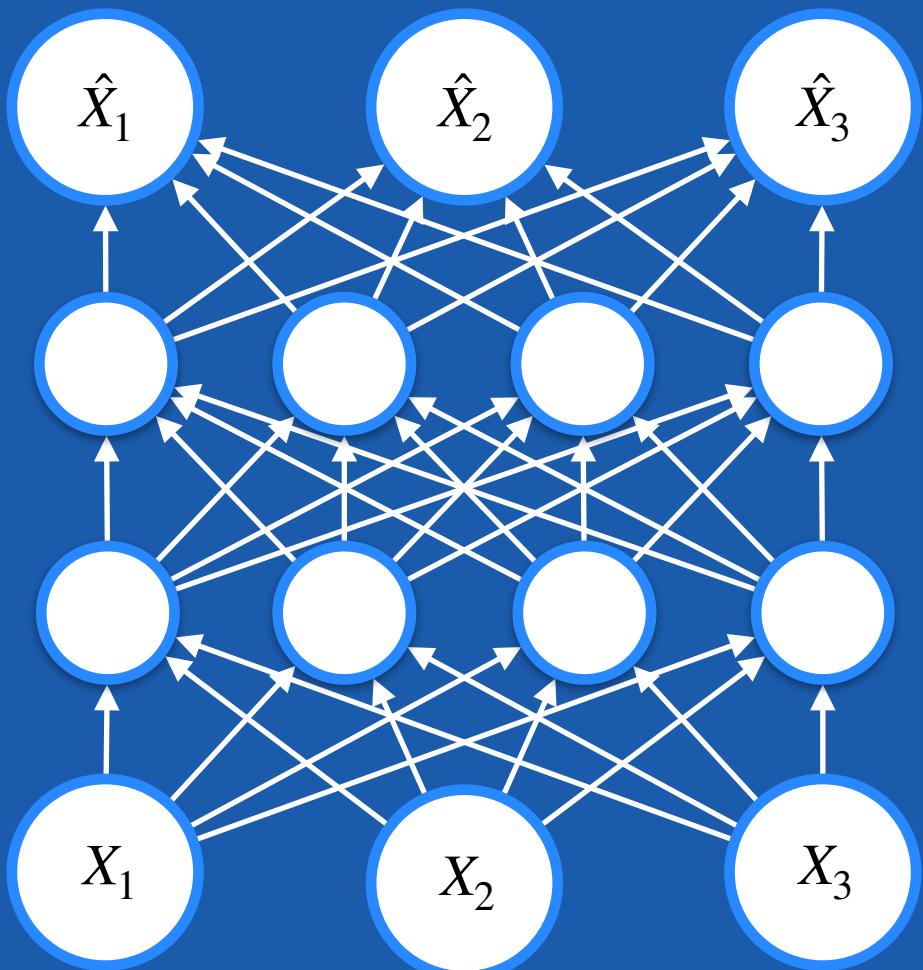
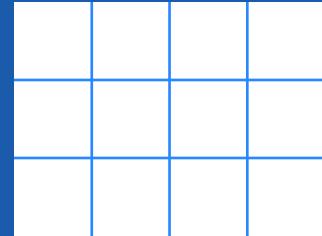
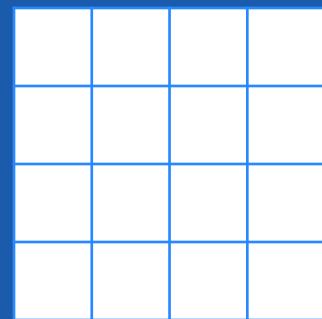
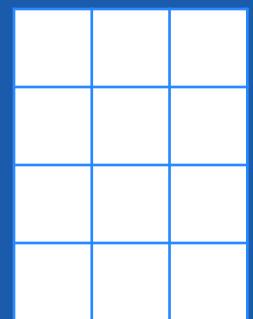
$$-\log p(x) = \sum_{d=1}^D -\log p(x_d | x_{<d})$$

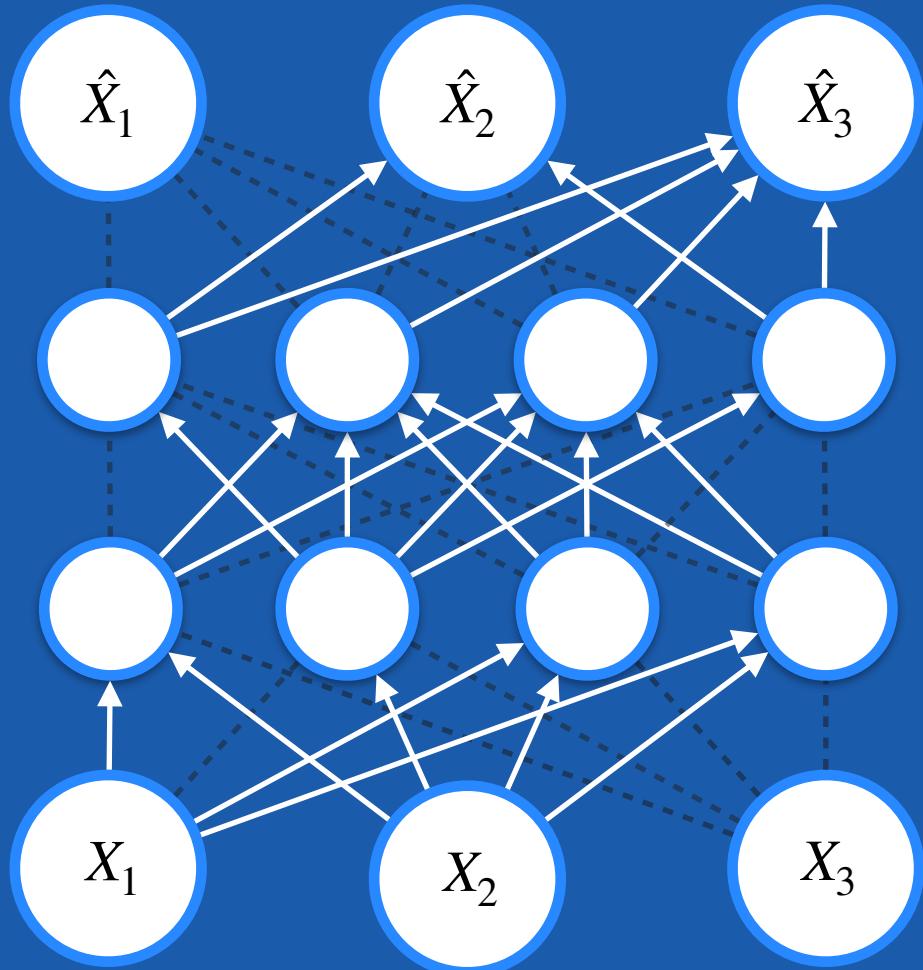
$$L(x) = \sum_{d=1}^D -x_d \log \hat{x}_d - (1 - x_d)(1 - \log \hat{x}_d)$$

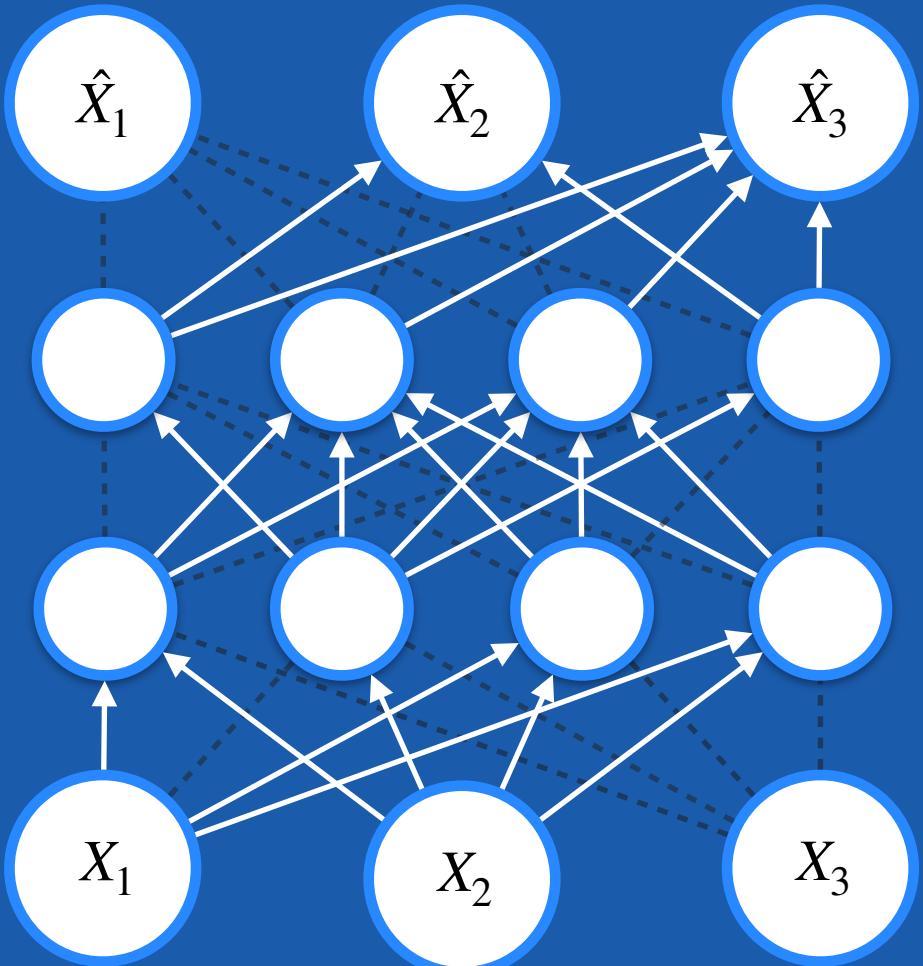
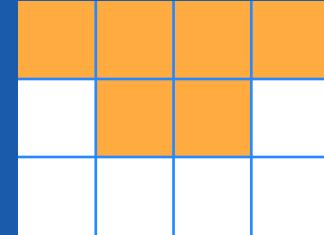
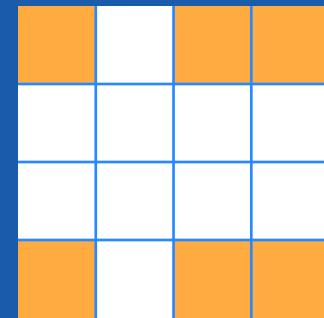
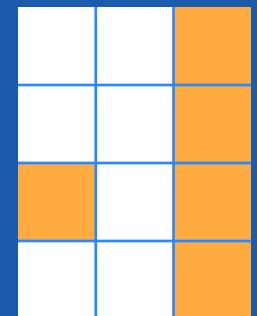


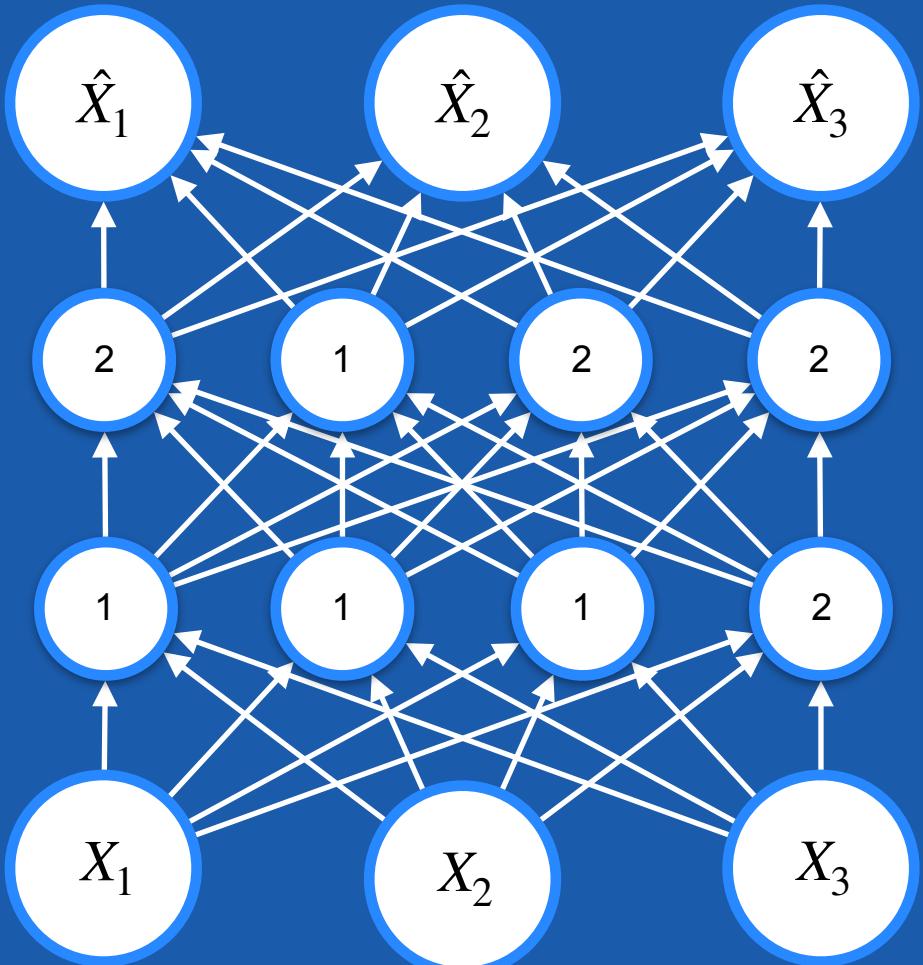
$$\hat{X} = \text{sigm}(\dots)$$

$$L(x) = \sum_{d=1}^D -x_d \log \hat{x}_d - (1 - x_d)(1 - \log \hat{x}_d)$$

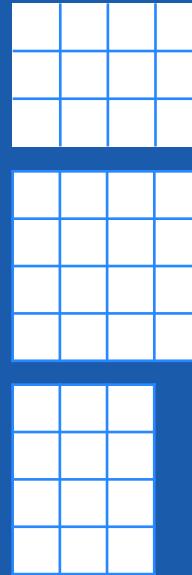

$$W_3$$

$$W_2$$

$$W_1$$




 M_3  M_2  M_1 

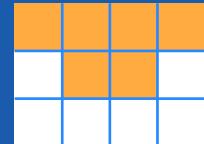


W_1



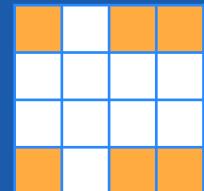
W_3

• M_3



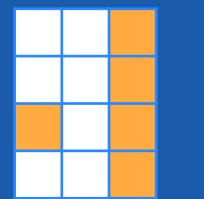
W_2

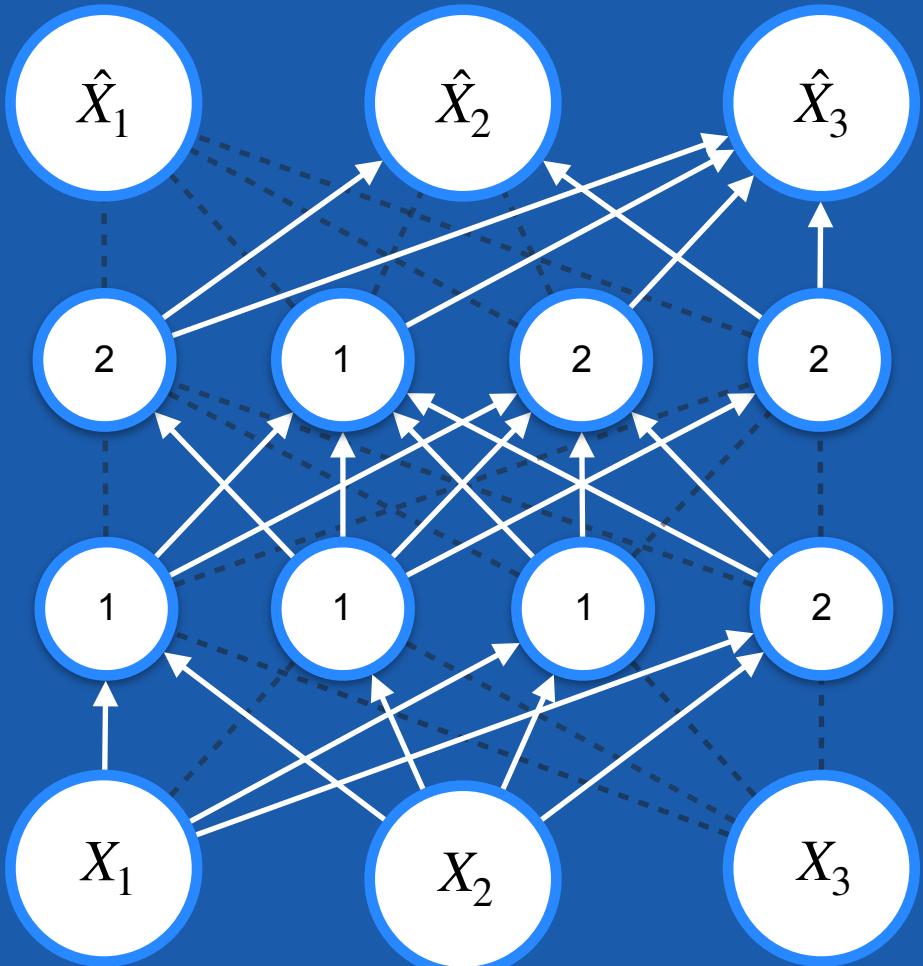
• M_2



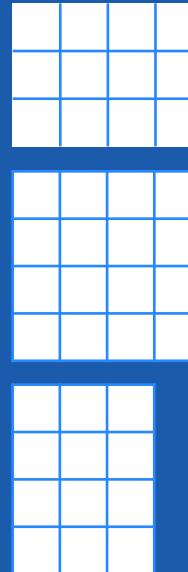
W_1

• M_1





W_1

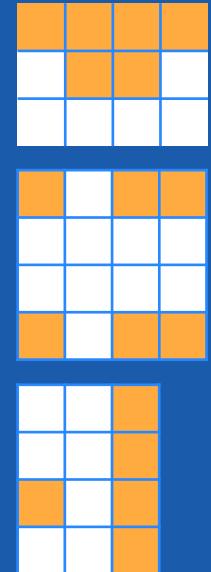


W_3

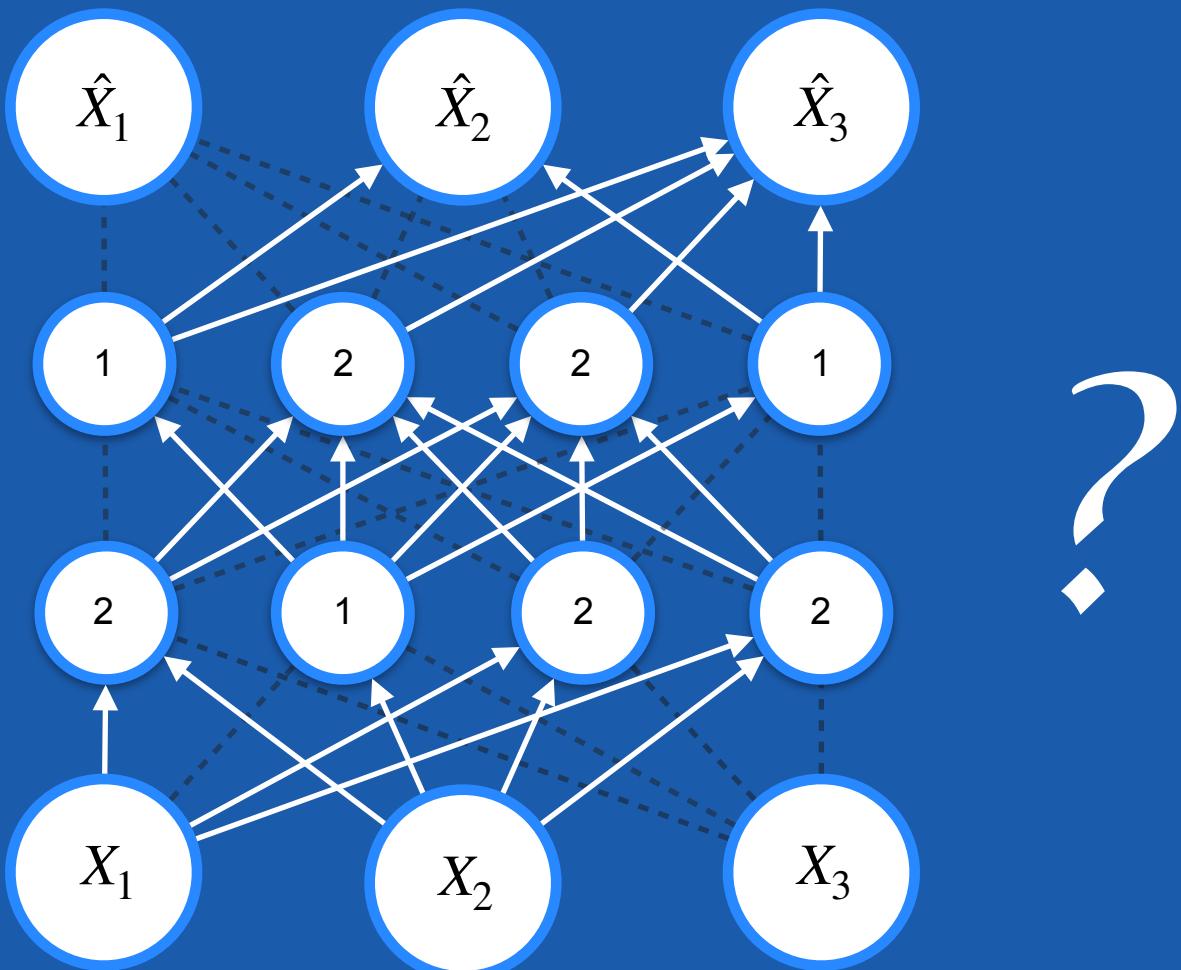
W_2

W_1

- M_3
- M_2
- M_1



Fin?



$$p(x) = p(x_3 | x_1, x_2)p(x_2 | x_1)p(x_1)$$

?

$$p(x) = p(x_3 | x_1, x_2)p(x_2 | x_1)p(x_1) \quad ?$$

$$p(x) = p(x_3 | x_2, x_1)p(x_1 | x_2)p(x_2) \quad ?$$

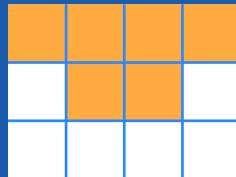
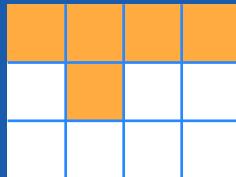
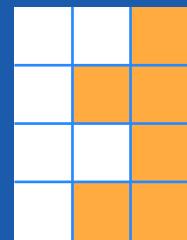
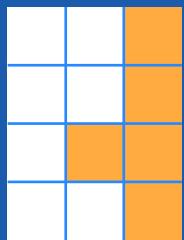
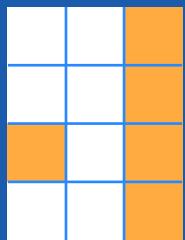
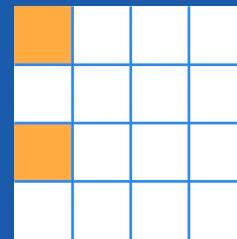
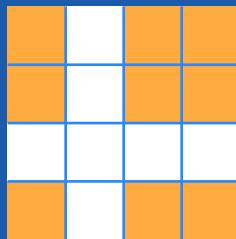
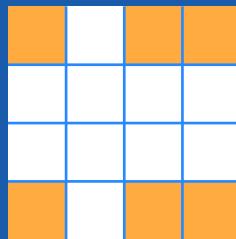
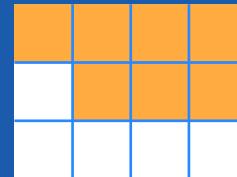
$$p(x) = p(x_2 | x_3, x_1)p(x_3 | x_1)p(x_1) \quad ?$$

$$p(x) = p(x_2 | x_1, x_3)p(x_1 | x_3)p(x_3) \quad ?$$

$$p(x) = p(x_1 | x_3, x_2)p(x_3 | x_2)p(x_2) \quad ?$$

$$p(x) = p(x_1 | x_2, x_3)p(x_2 | x_3)p(x_3) \quad ?$$

- ◆ Order Agnostic training
- ◆ Connection Agnostic training

\mathcal{M}_1  \mathcal{M}_2  \mathcal{M}_3 

QA