Training and Inference on Any-Order Autoregressive Models the Right Way

Andy Shih, Dorsa Sadigh, Stefano Ermon

andyshih@cs.stanford.edu

Selected As Oral

Joint Likelihood

Autoregressive Model

$$\log p(\boldsymbol{x}) = \sum_{i=1}^{N} \log p(x_i | \boldsymbol{x}_{<i})$$

joint — univariate conditional

Marginal Likelihood

Any-Order Autoregressive Model

$$\log p(\boldsymbol{x}_e)$$

marginal

Image

Language

I l_e _o _l_y s__rts

Planning

Pertsch et al.
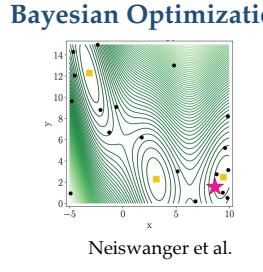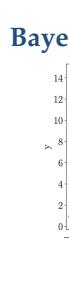
Bayesian Optimization

Neiswanger et al.

**Probabilistic Programming**

Wu et al.

and more…

Difficulty: The evidence subset is different for each query!

$$p(x_1, x_3) = \int_{x_2} \int_{x_4} p(x_1, x_2, x_3, x_4) dx_2 dx_4$$

high-dimensional integration over missing variables ☹

💡 compute marginals via chain rule!

$p(x_1)$ $p(x_2|x_1)$ $p(x_3|x_1,x_2)$ $p(x_4|x_1,x_2,x_3)$

$p(x_4)$ $p(x_3|x_4)$ $p(x_2|x_3,x_4)$ $p(x_1|x_2,x_3,x_4)$
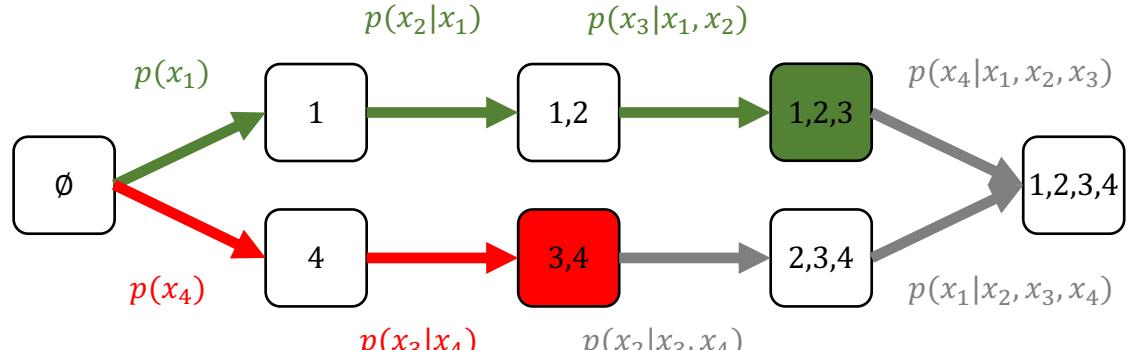
$$\log p(\boldsymbol{x}_{1:3}) = \log p(x_1) + \log p(x_2|x_1) + \log p(x_3|x_1,x_2)$$

$$\log p(\boldsymbol{x}_{3:4}) = \log p(x_4) + \log p(x_3|x_4)$$
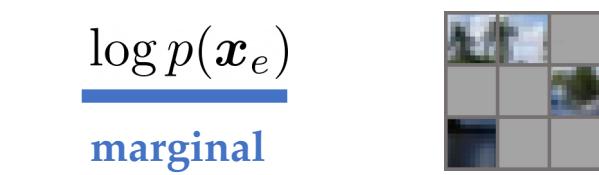
Any-Order Autoregressive Models learn *all* orderings

$$p(x_3|x_2, x_4)$$
$$p(x_3|x_1)$$
$$p(x_2|x_{\emptyset})$$

All univariate conditionals (edges) learned with a **single** weight-tied neural network

Input $\boldsymbol{x}$ →

Mask e →

model → $p(x_i|\boldsymbol{x}_e)$
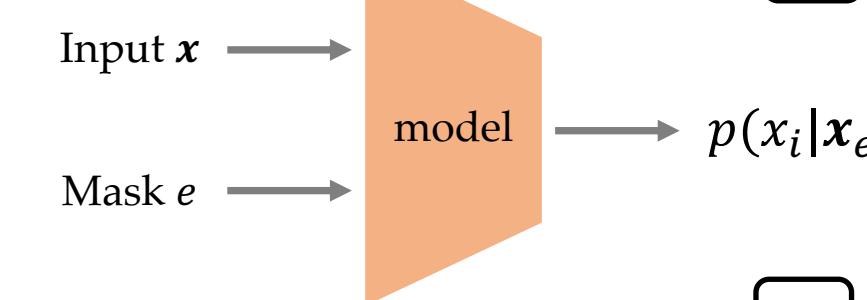
MAC: Mask-Tuned Arbitrary Conditional Models

Two improvements:

1   For each node, we only need one path from $\emptyset$

    Only need the red edges!

2   Some edges will be evaluated more frequently!
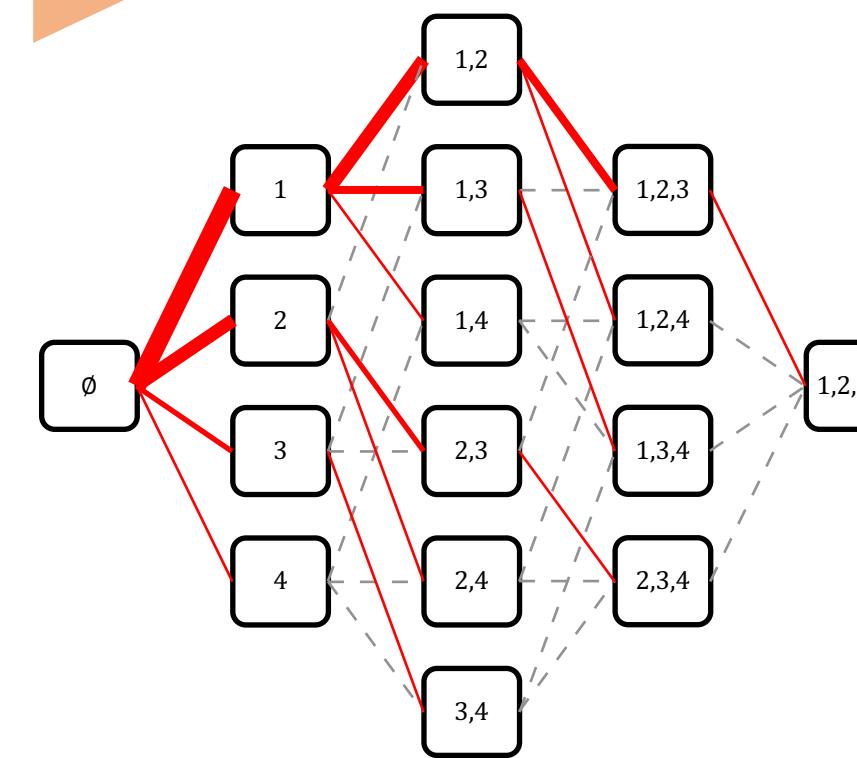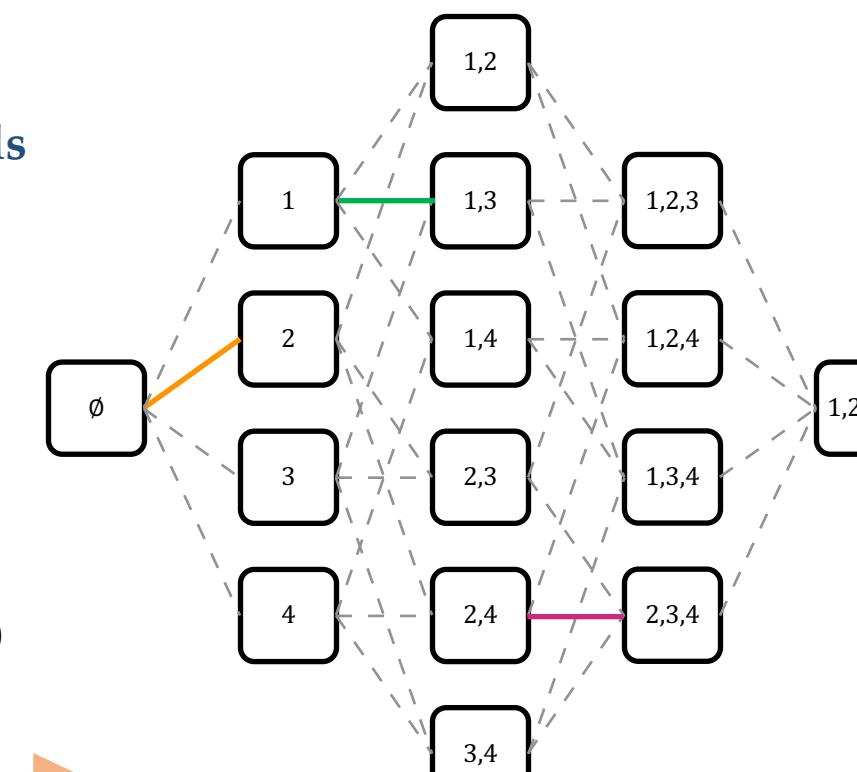
Text8 dataset (bpd, lower is better)

	joint	marginal
OA-Transformer	1.64	
D3PM	1.47	
ARDM (14000 epochs)	1.43	
ARDM (3000 epochs)	1.48	1.12
MAC (3000 epochs)	**1.40**	**1.09**
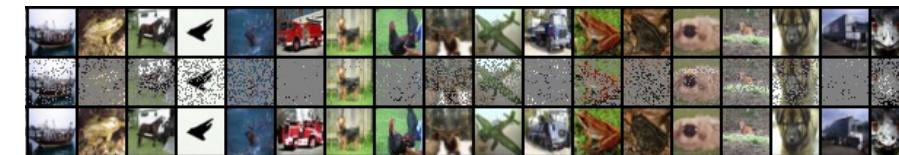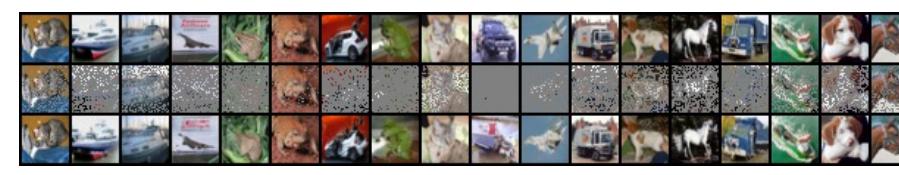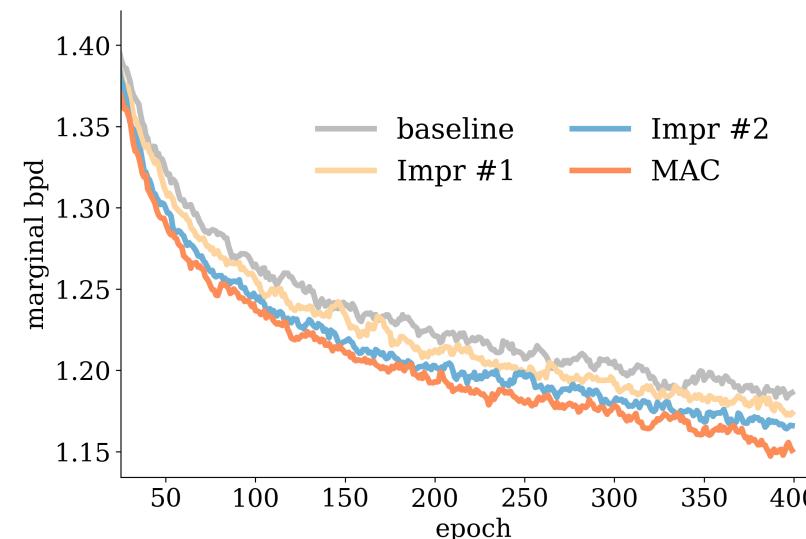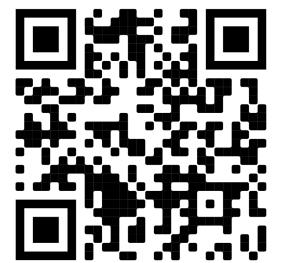
legend: baseline, Impr #1, Impr #2, MAC


```
prophylactic drugs several drugs most of which are also used for trea
tment of malaria can be taken preventatively generally these drugs are
taken daily or weekly at a lower dose than would be used for treatmen
t of a person who had actually contracte

pr_p_y__cti__dr__s _eve__l drug_ _o_t of__h__h are __so u_ed __r _re_
tment__f mal__i_c__ b_ _a_en _re_enta_vely ge__ra_l_ t_es_ d_ugs_are
_ake__aily or _ee_ly _t a lo_er dose t__n _o_ld _e_used_f_r _re_tme_
t__f__ _e__on_w____ad_a_tuall___on_ra_te

prophylactic drugs several drugs most of which are also used for trea
tment of malaria can be taken preventatively generally these drugs are
taken daily or weekly at a lower dose than would be used for treatmen
t of a lesion who had actually contracte
```