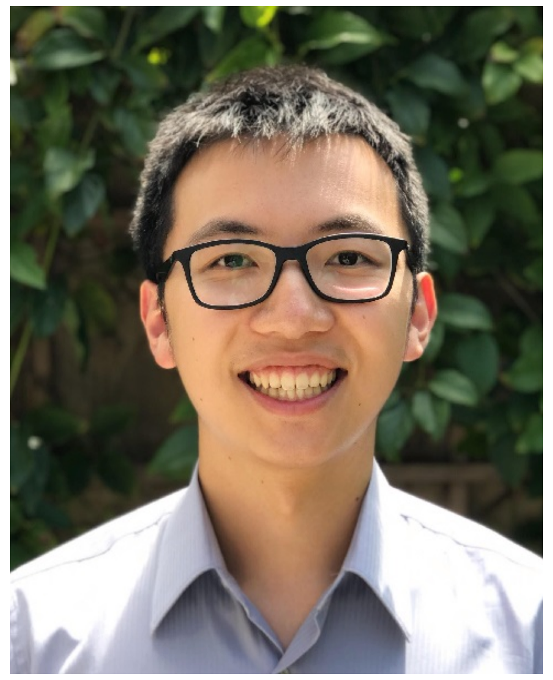
HyperSPNs: Compact and Expressive Probabilistic Circuits



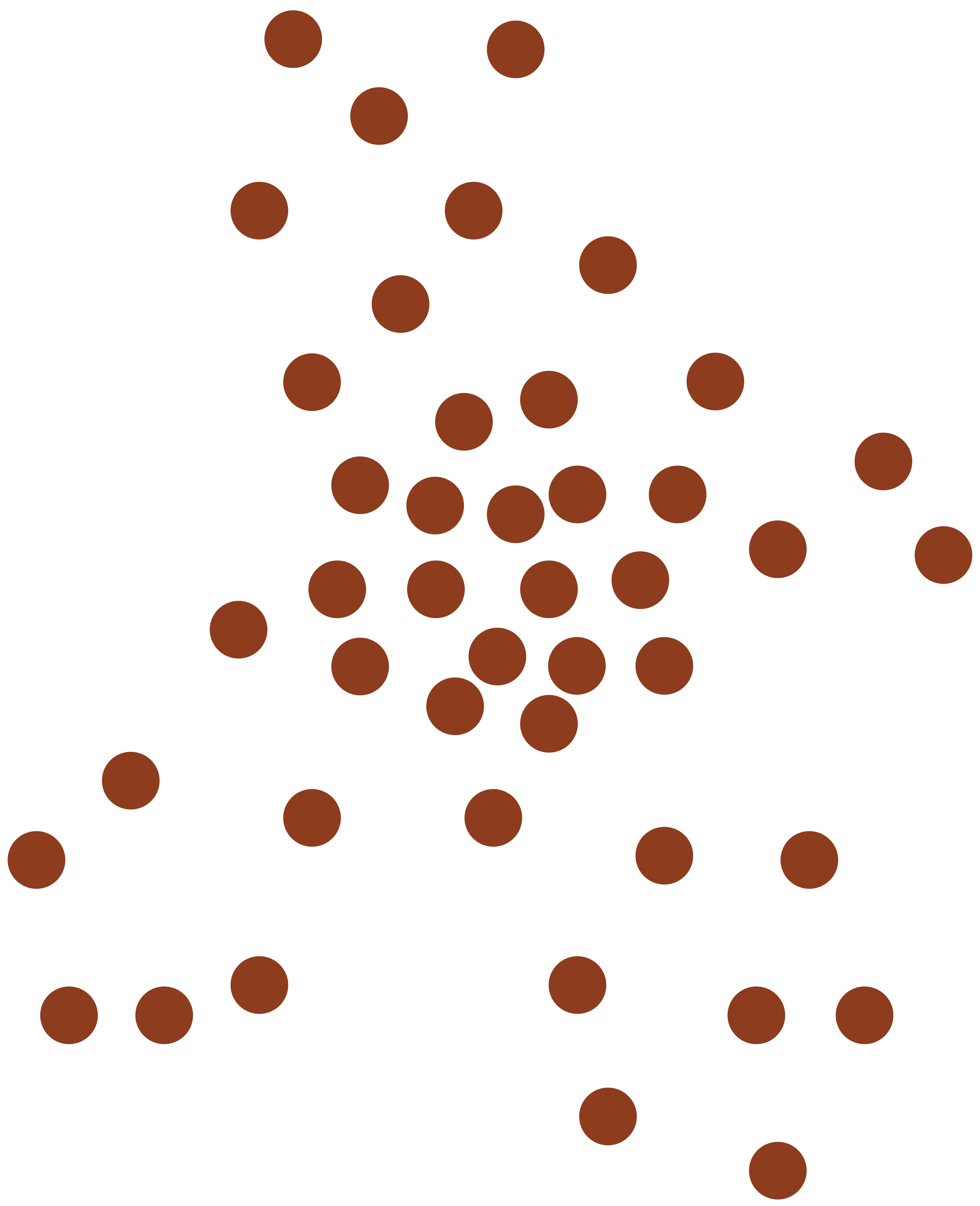
Andy Shih

Dorsa Sadigh

Stefano Ermon

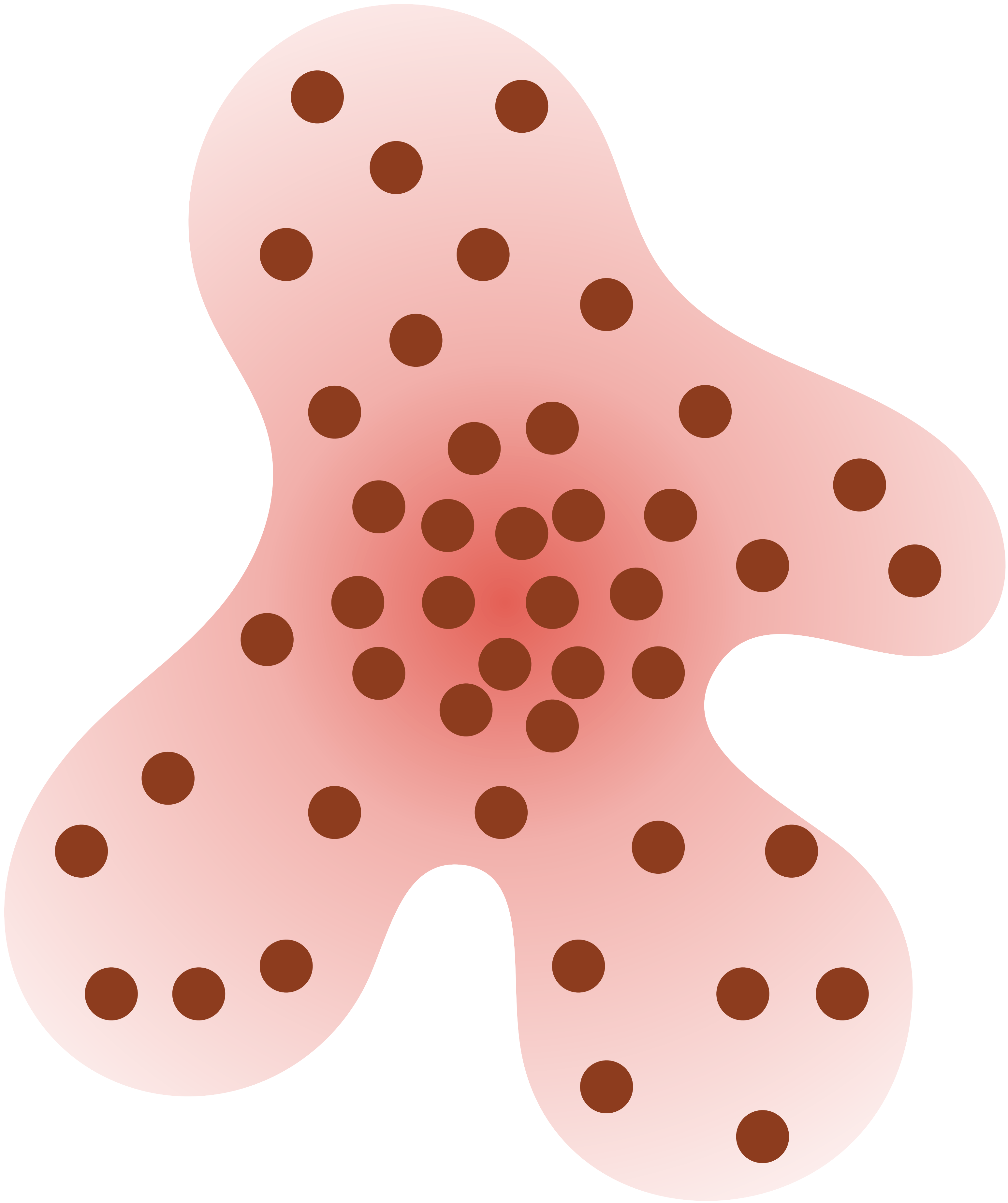
Stanford University

Density Estimation



Data

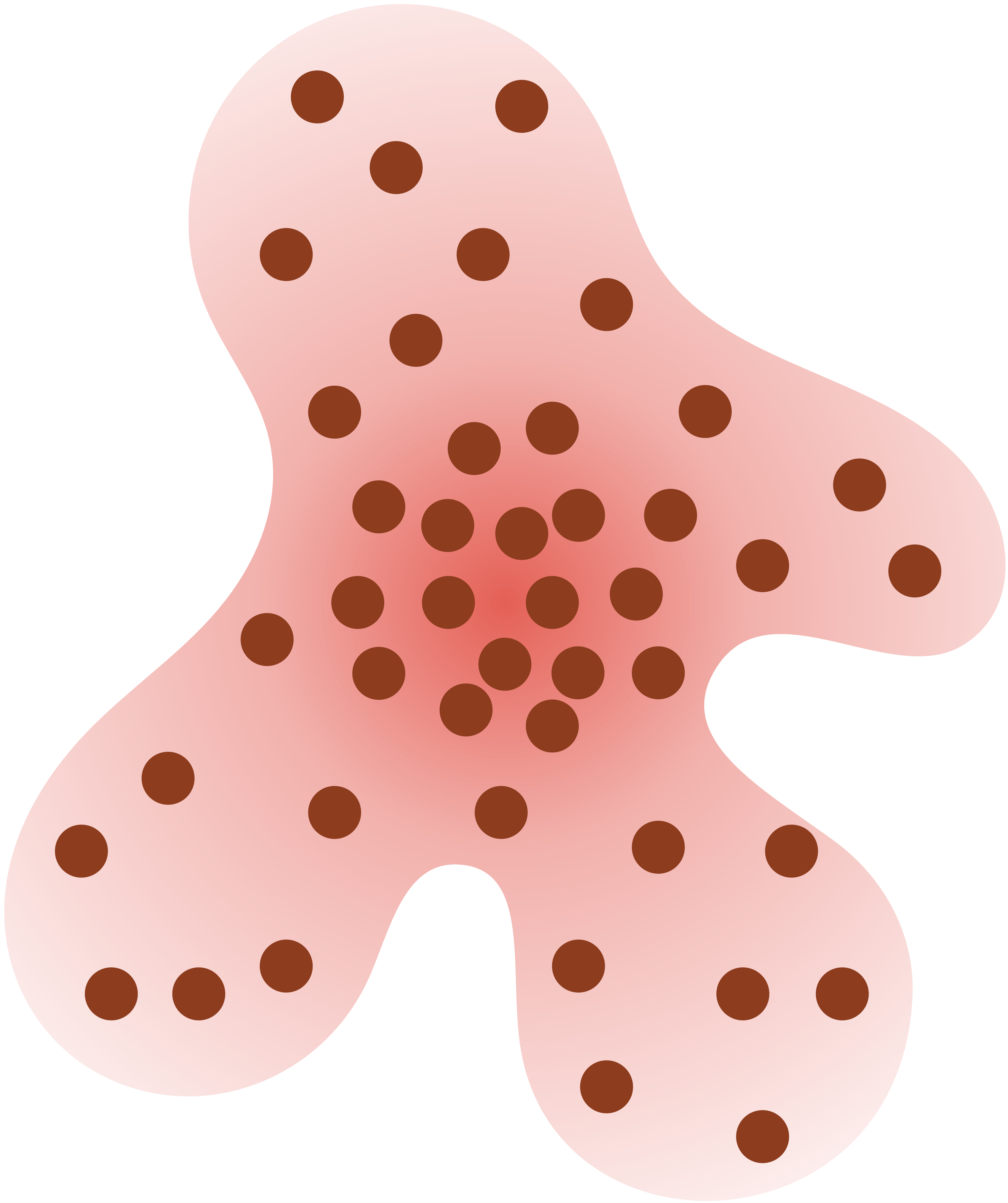
Density Estimation



Data

Distribution

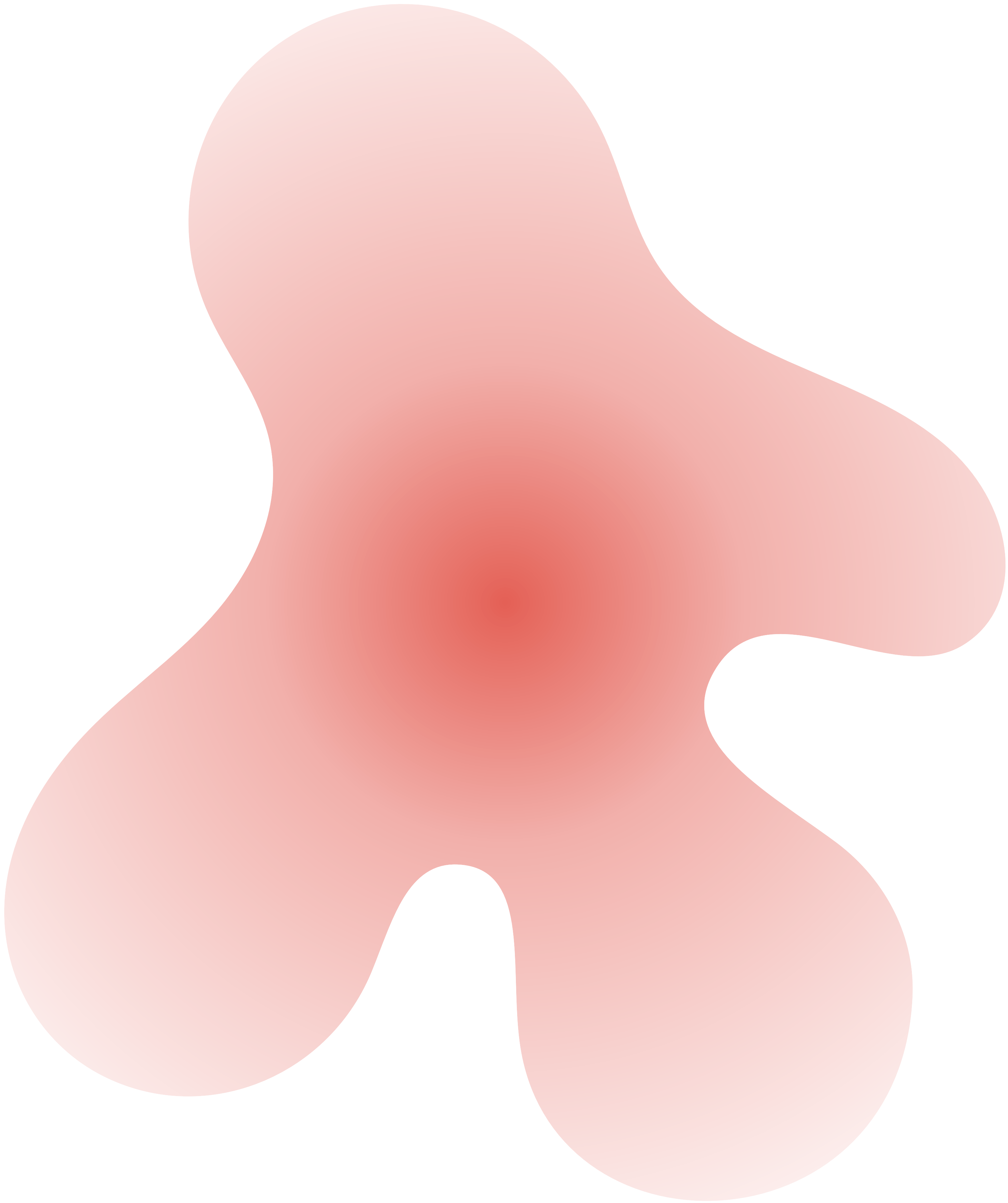
Density Estimation



~~Data~~

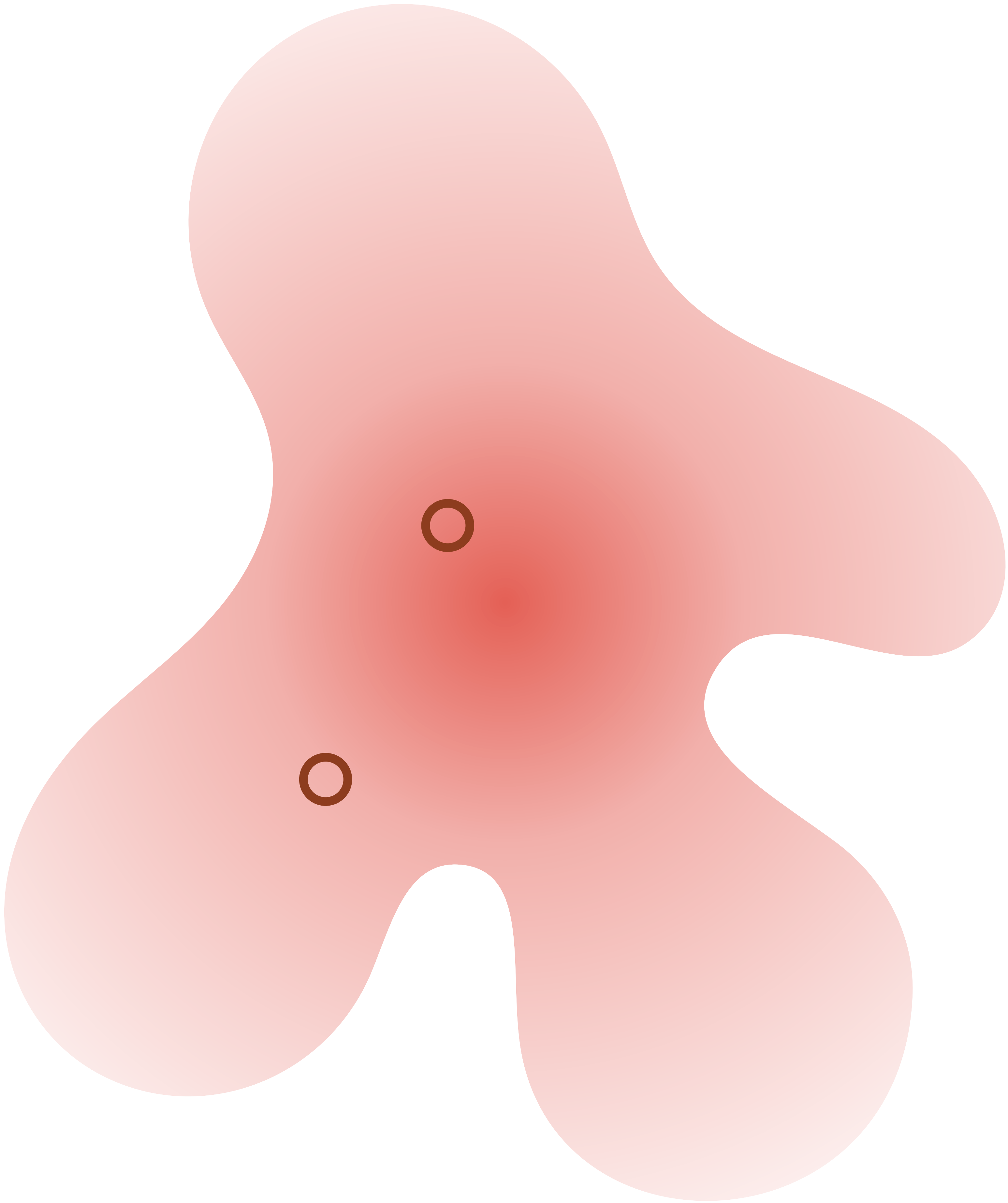
Distribution

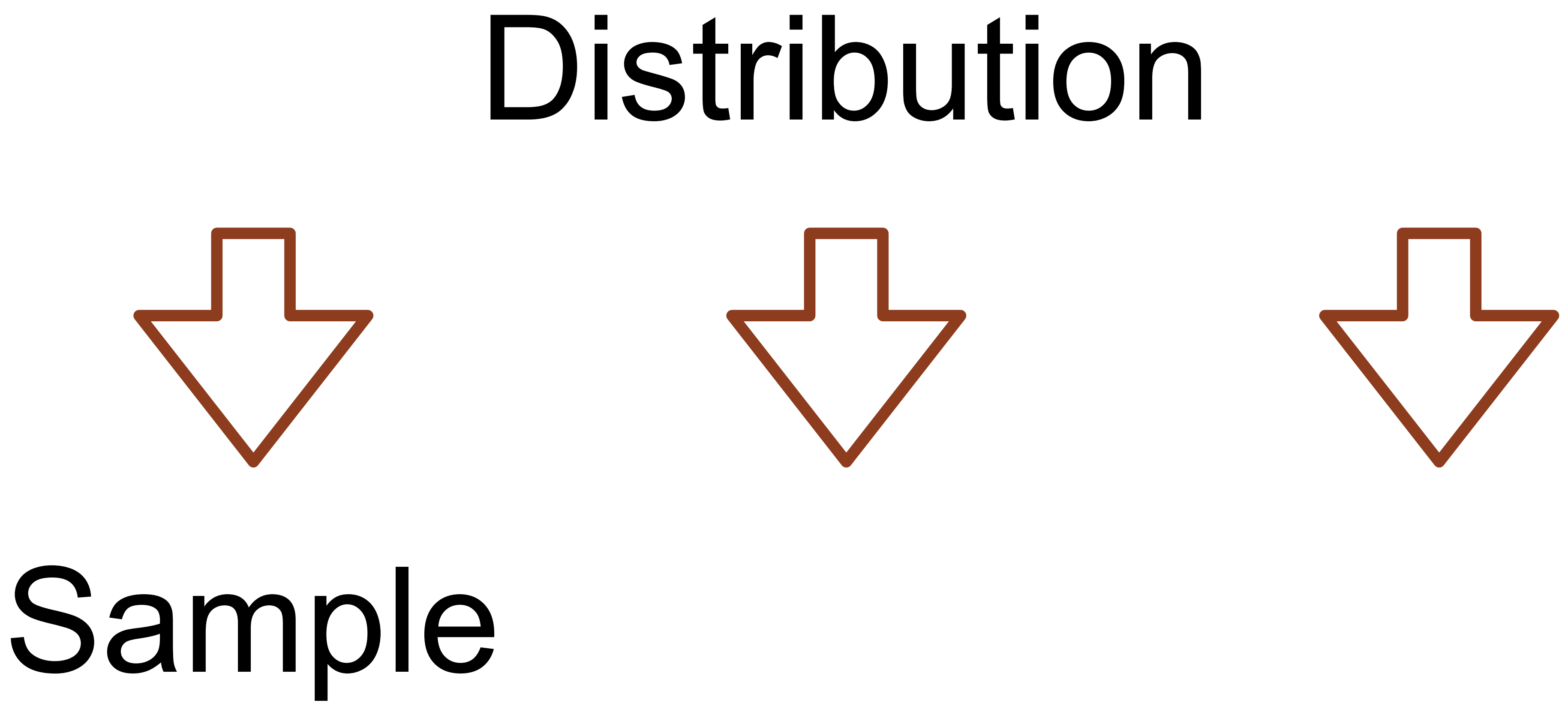
Density Estimation



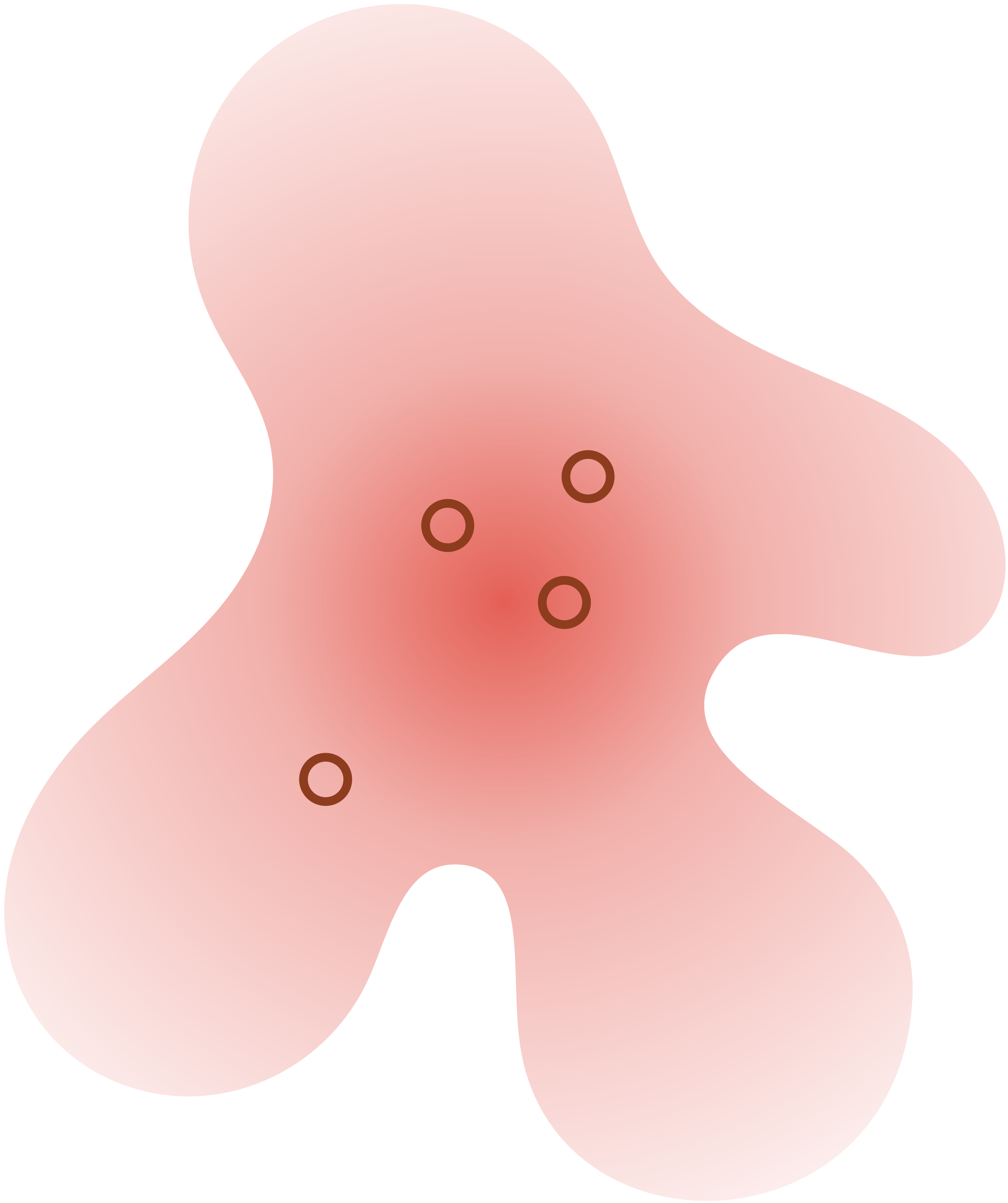
Distribution

Density Estimation





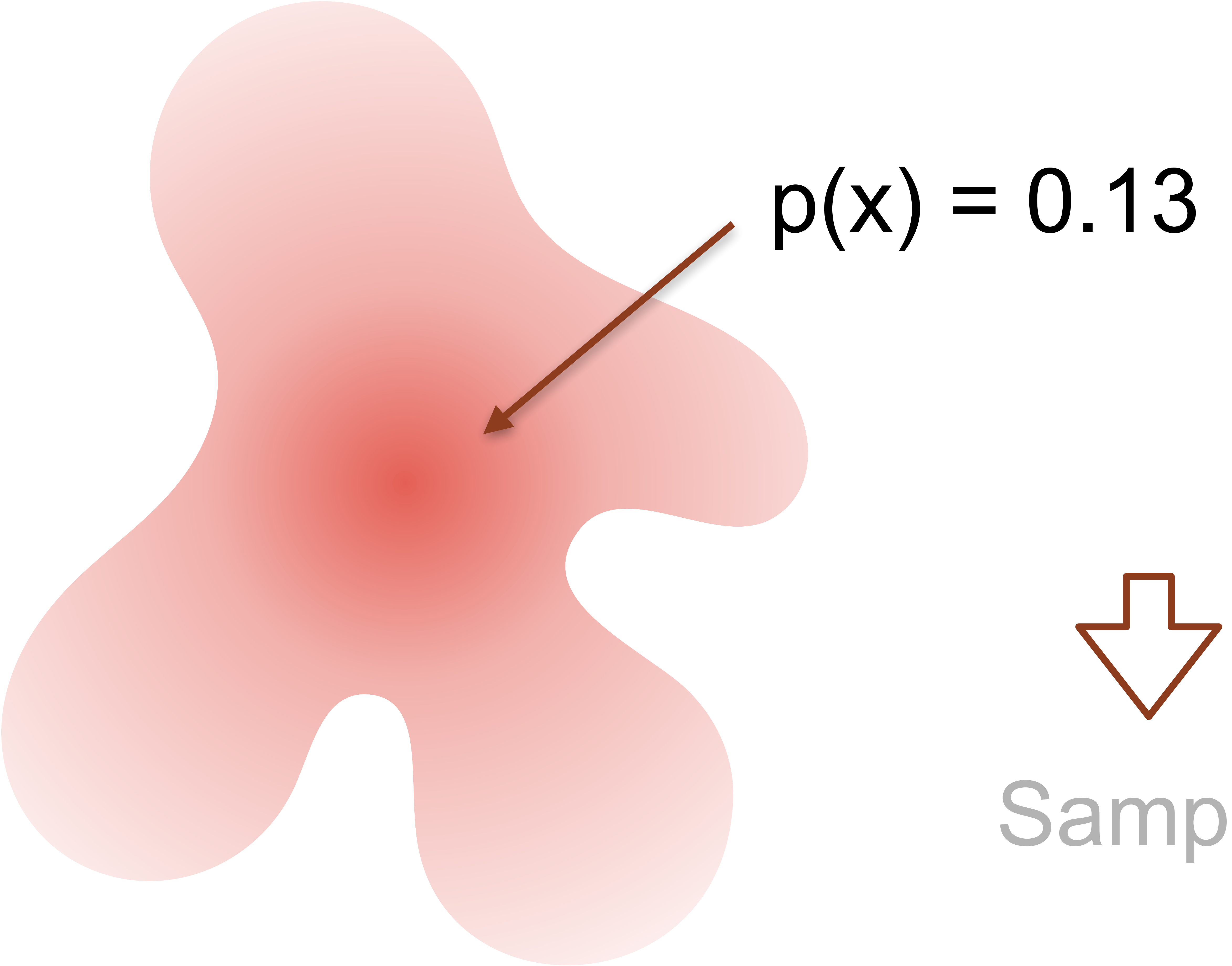
Density Estimation



Distribution

Sample

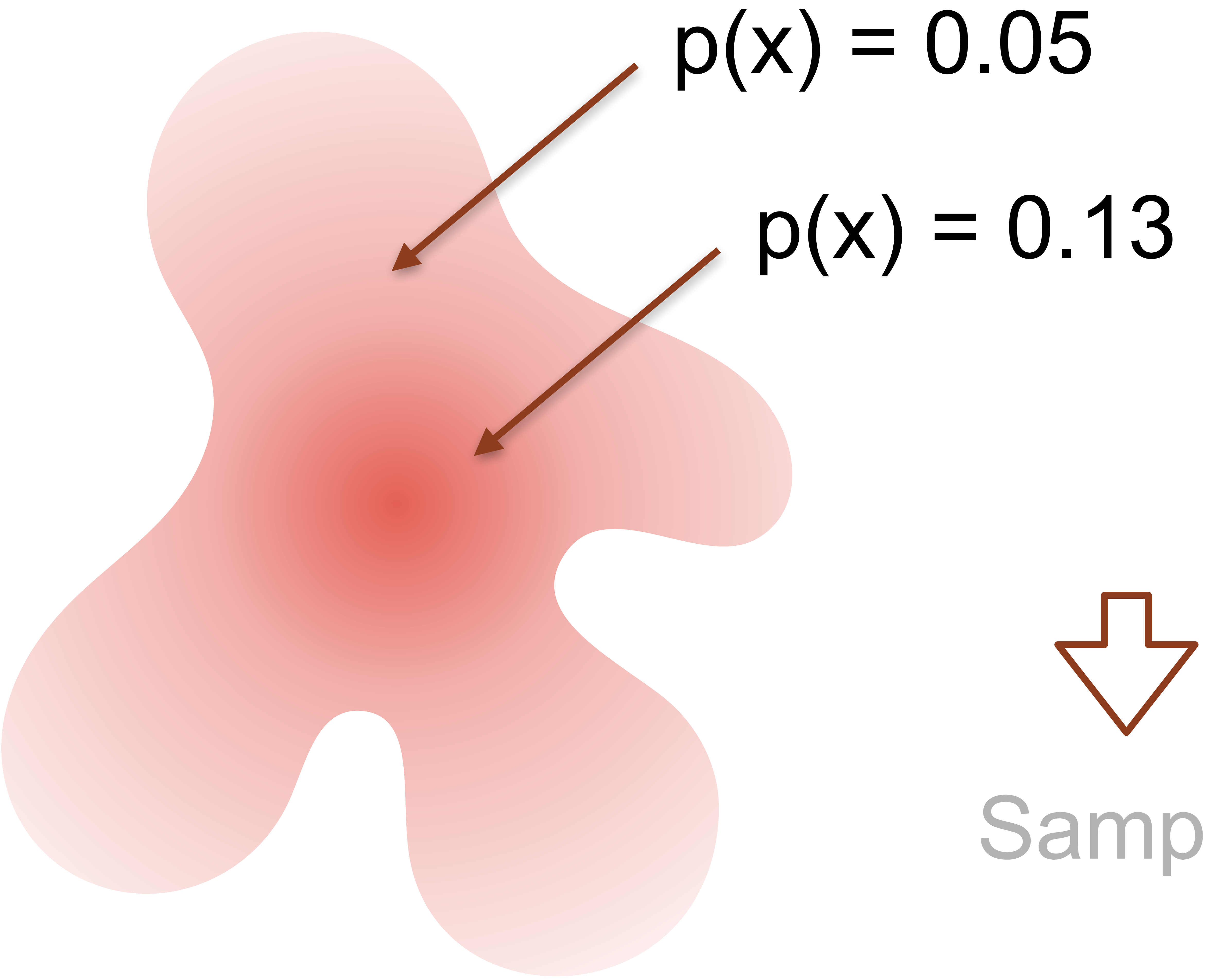
Density Estimation



Distribution

Sample Density

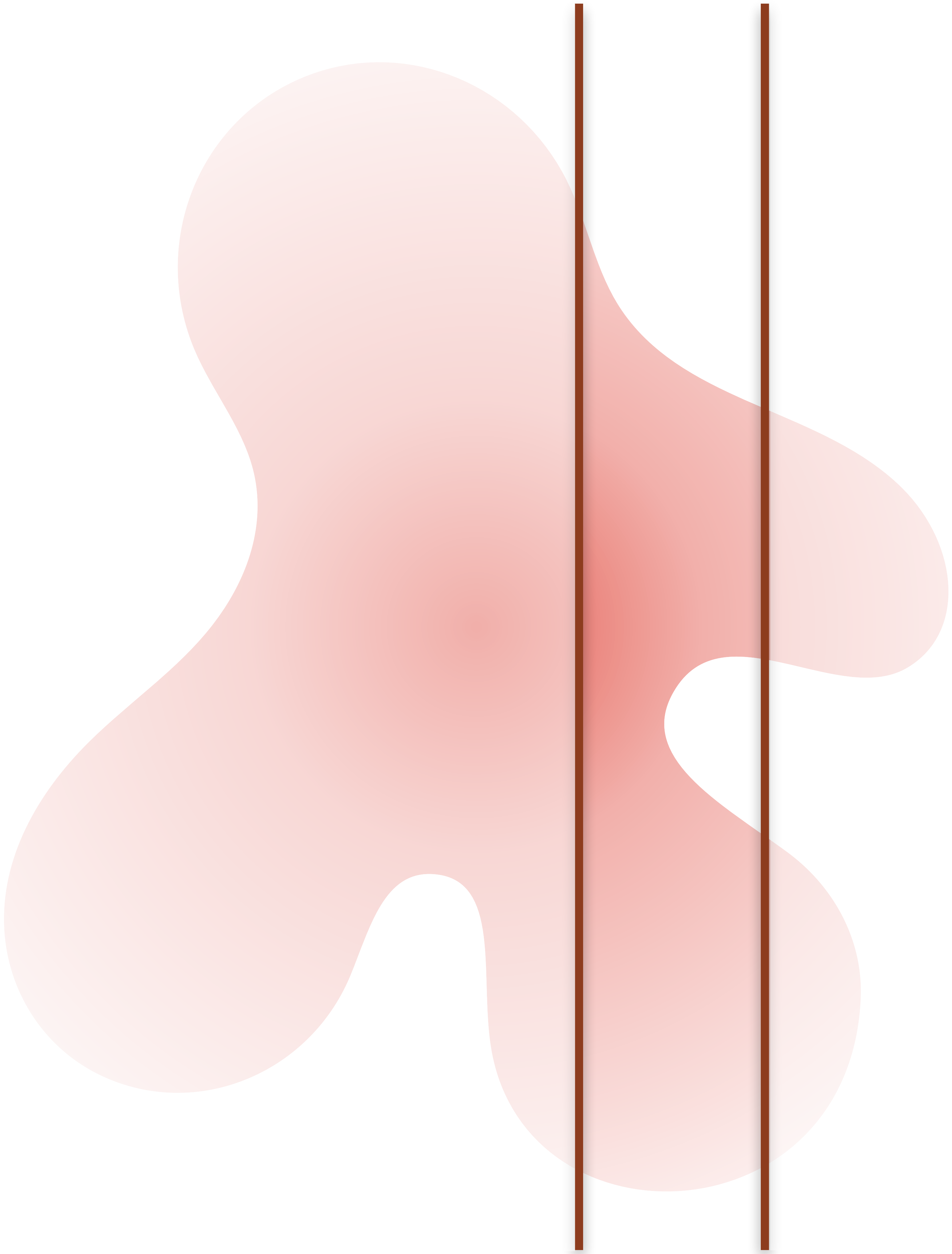
Density Estimation



Distribution

Sample Density

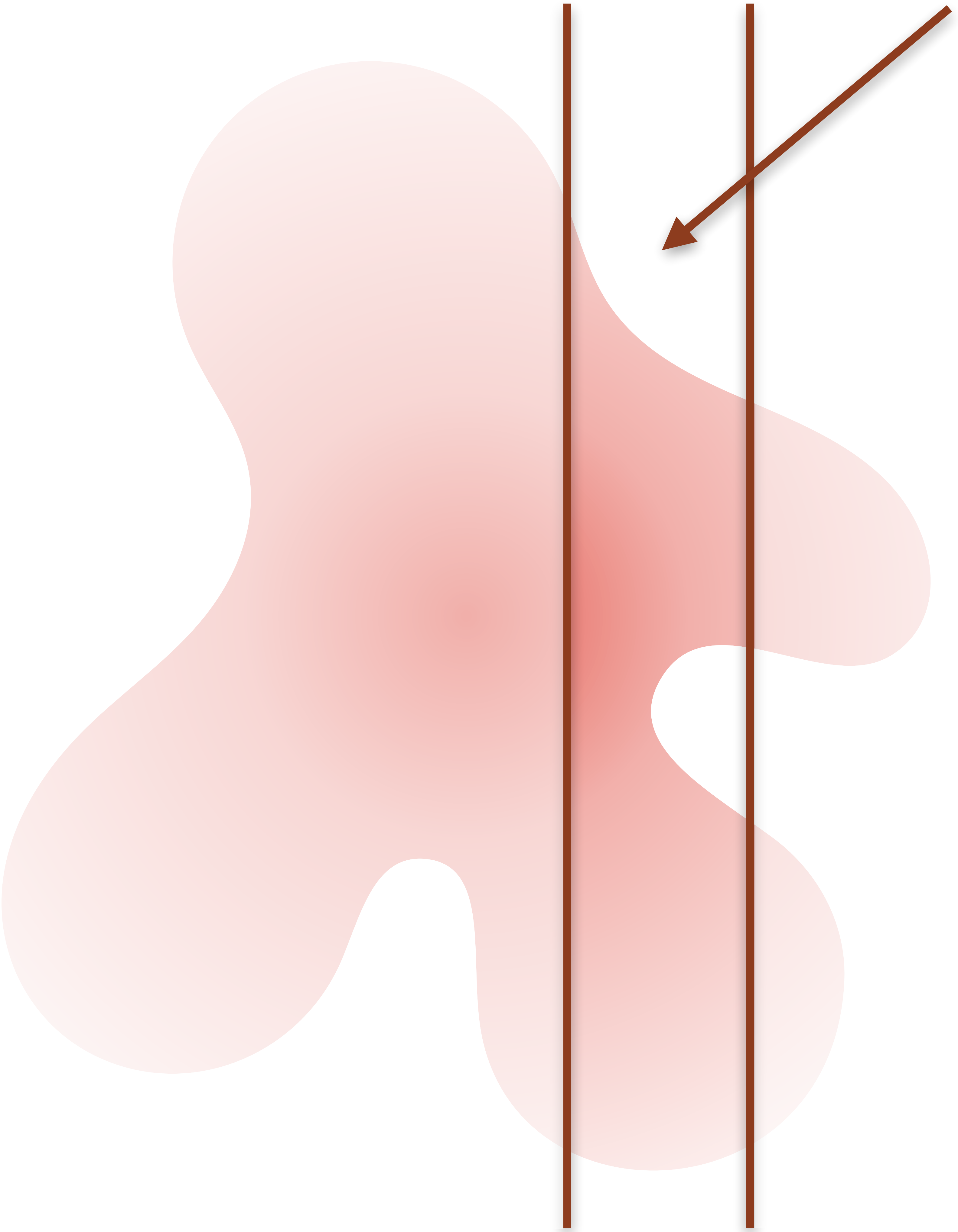
Density Estimation



Distribution

Sample Density Inference?

Density Estimation



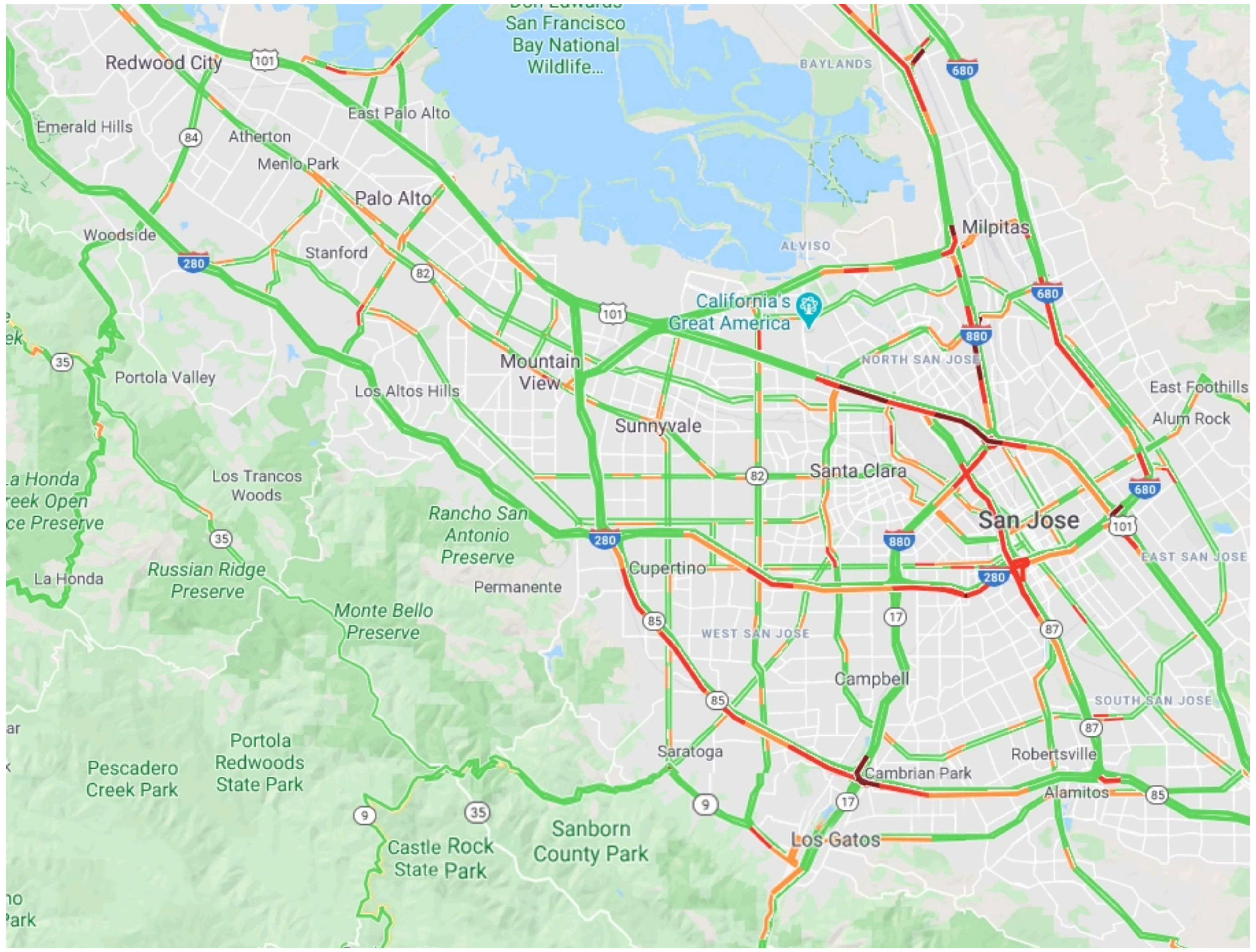
p( ) = ?

Distribution

↓ ↓ ↓

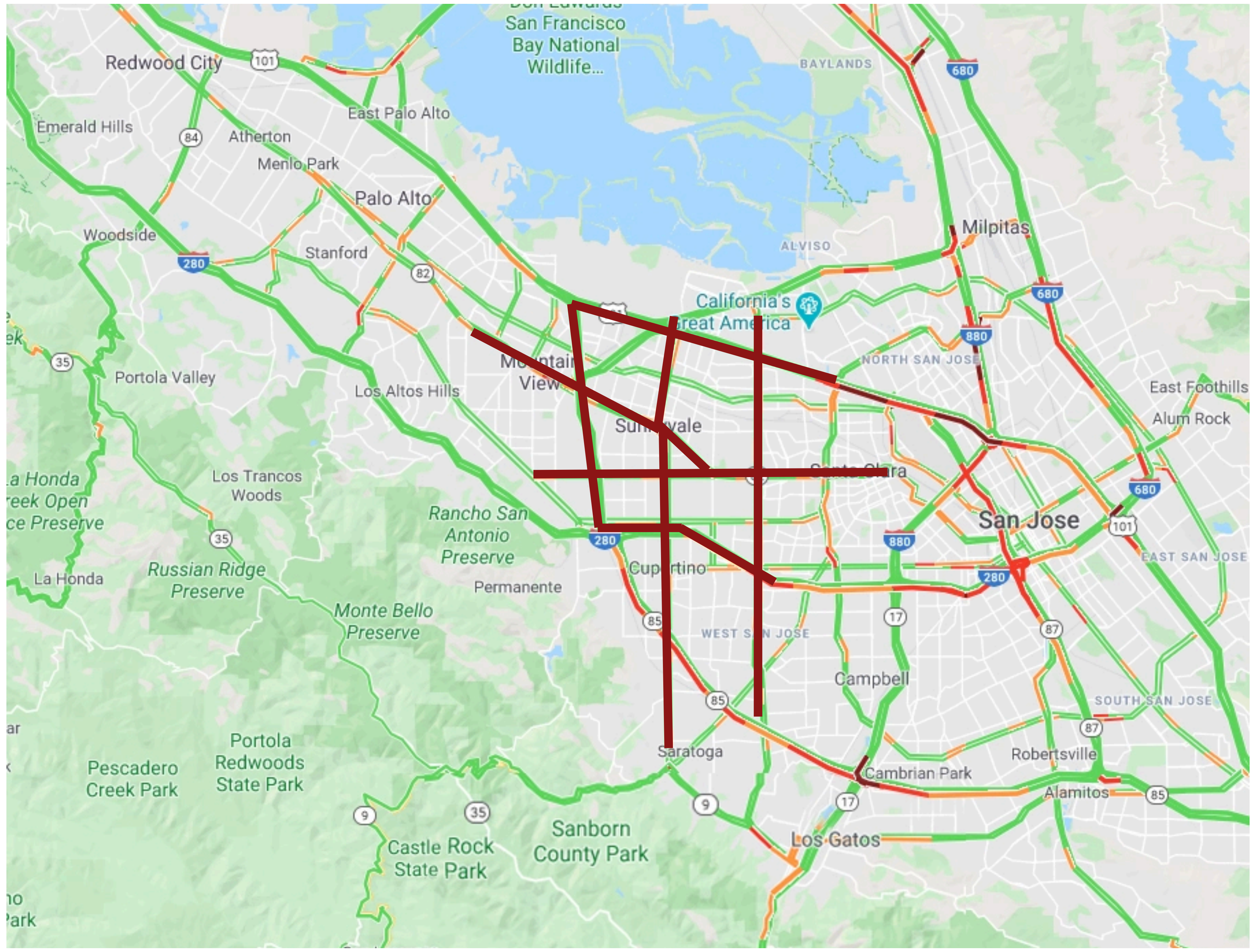
Sample Density Inference?

Inference - Marginals & Conditionals



Inference - Marginals & Conditionals

$$X = \{r_1, r_2, \ldots, r_{100}\}$$

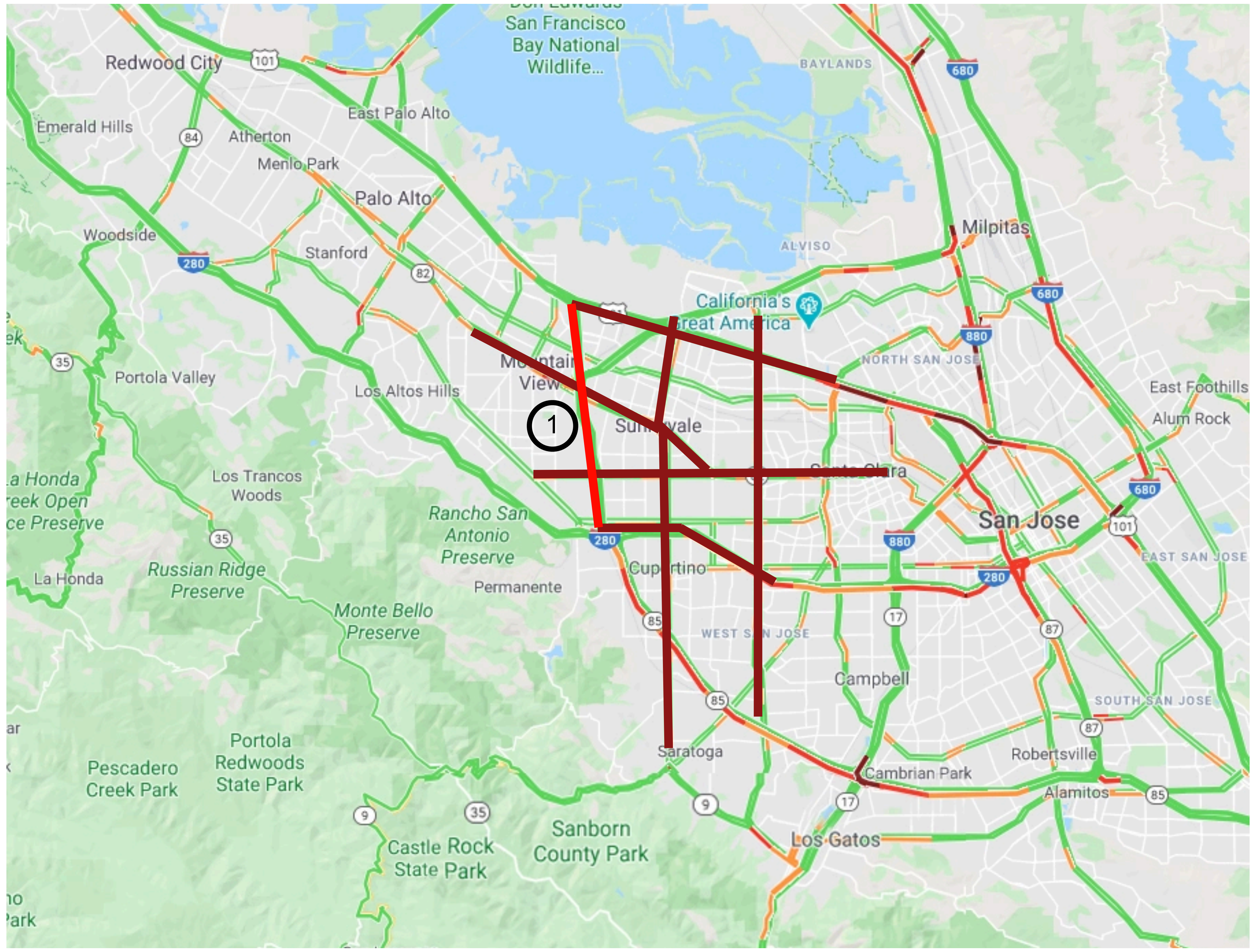


Inference - Marginals & Conditionals

$$X = \{r_1, r_2, \ldots, r_{100}\}$$

What's the probability that:

- road 1 is under construction?



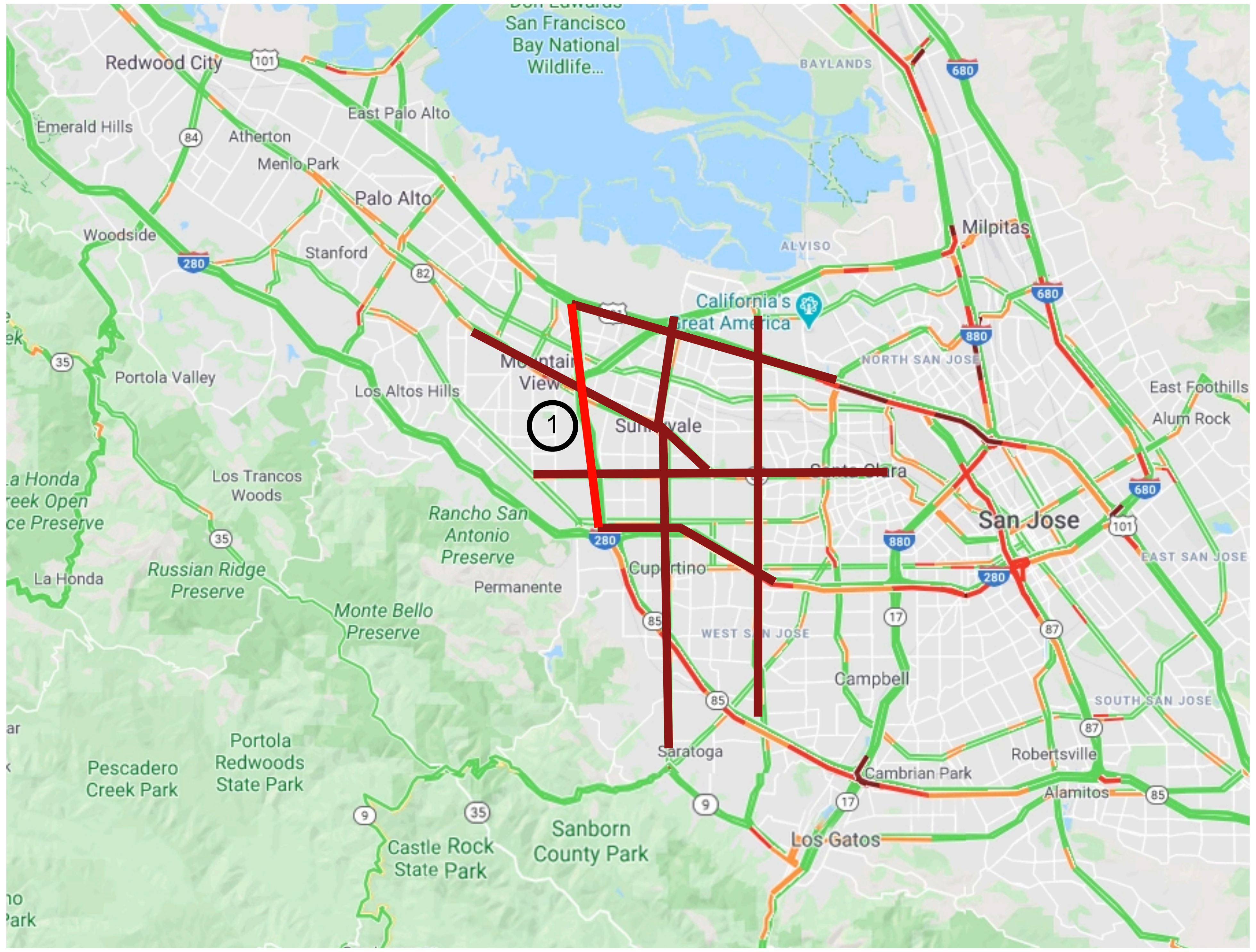
Inference - Marginals & Conditionals

$$X = \{r_1, r_2, \ldots, r_{100}\}$$

What's the probability that:

- road 1 is under construction?

$$\sum_{r_2,\ldots,r_{100}} p(r_1 = c, r_2, \ldots, r_{100})$$



Inference - Marginals & Conditionals

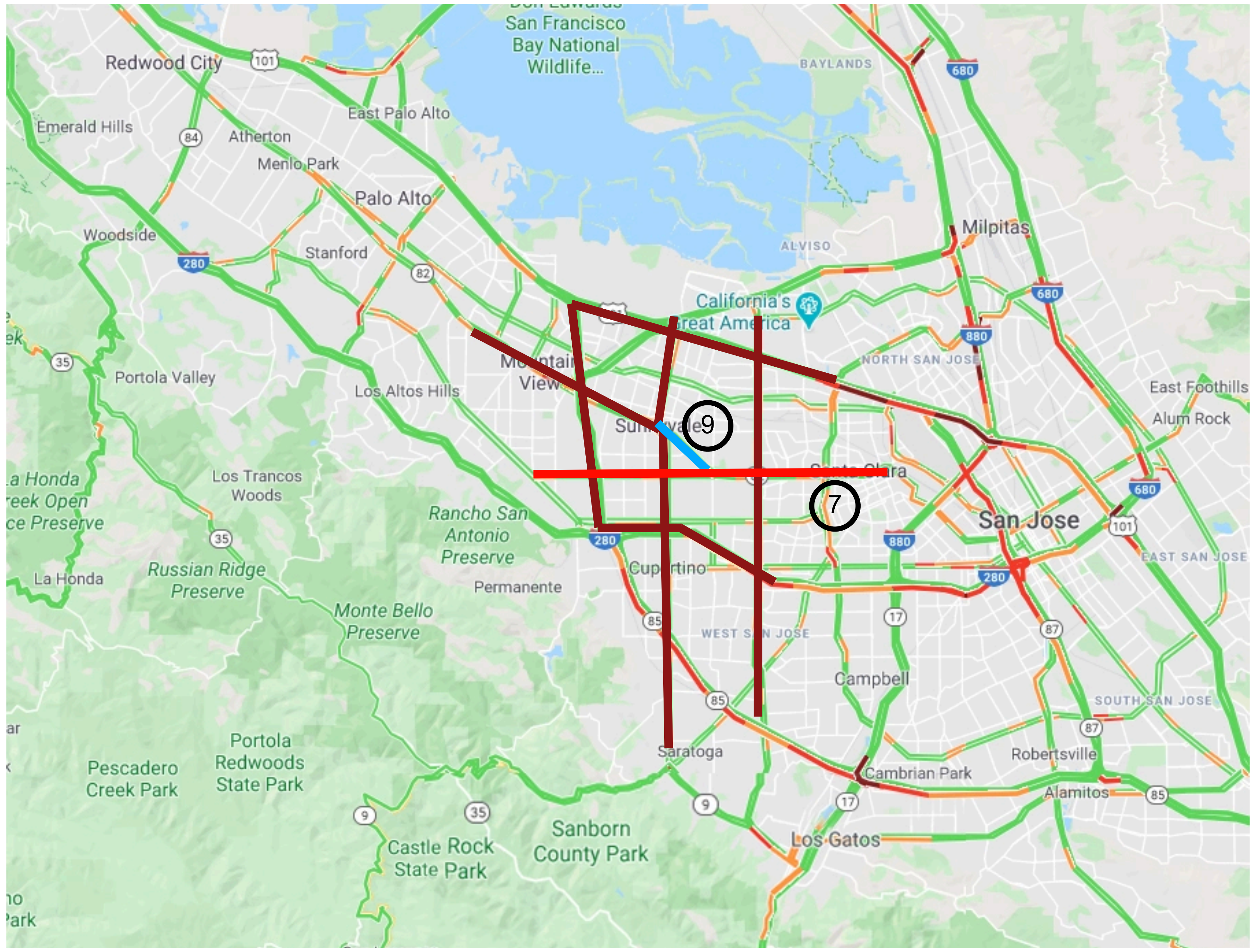
$$X = \{r_1, r_2, \ldots, r_{100}\}$$

What's the probability that:

- road 1 is under construction?

$$\sum_{r_2,\ldots,r_{100}} p(r_1 = c, r_2, \ldots, r_{100})$$

- road 7 is busy given road 9 is under construction?



Inference - Marginals & Conditionals

$$X = \{r_1, r_2, \ldots, r_{100}\}$$

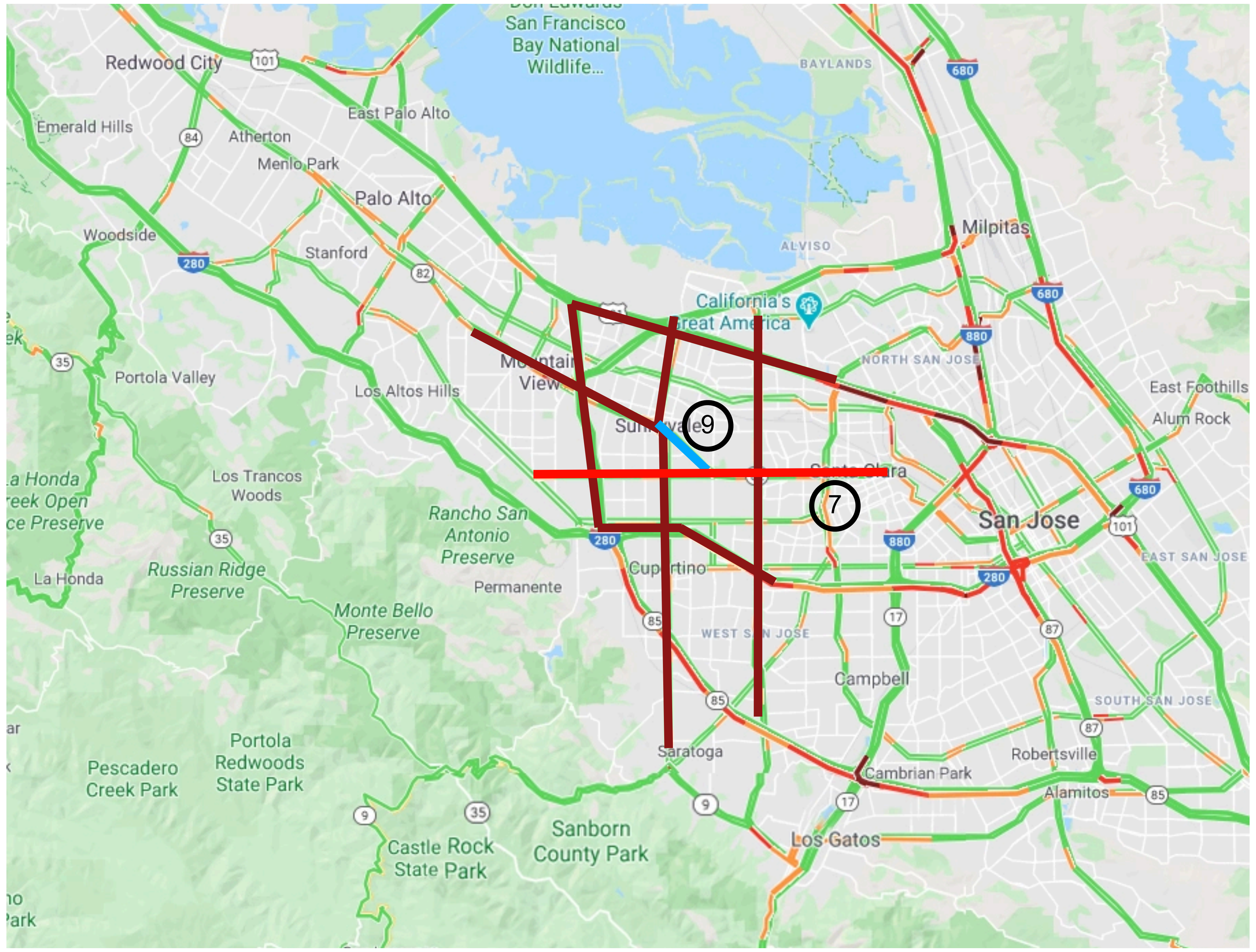
What's the probability that:

- road 1 is under construction?

$$\sum_{r_2,\ldots,r_{100}} p(r_1 = c, r_2, \ldots, r_{100})$$

- road 7 is busy given road 9 is under construction?

$$p(r_7 = b, r_9 = c)/p(r_9 = c)$$



Modeling Families

Mixture Models

tractable

expressive

Modeling Families

Mixture Models

tractable

~~expressive~~

Modeling Families

Mixture Models

tractable

expressive
efficient

Modeling Families

Score-based

VAEs

Flows

GANs

Mixture Models

tractable

expressive
efficient

Modeling Families

Score-based

VAEs

Flows

GANs

DPPs

SPNs

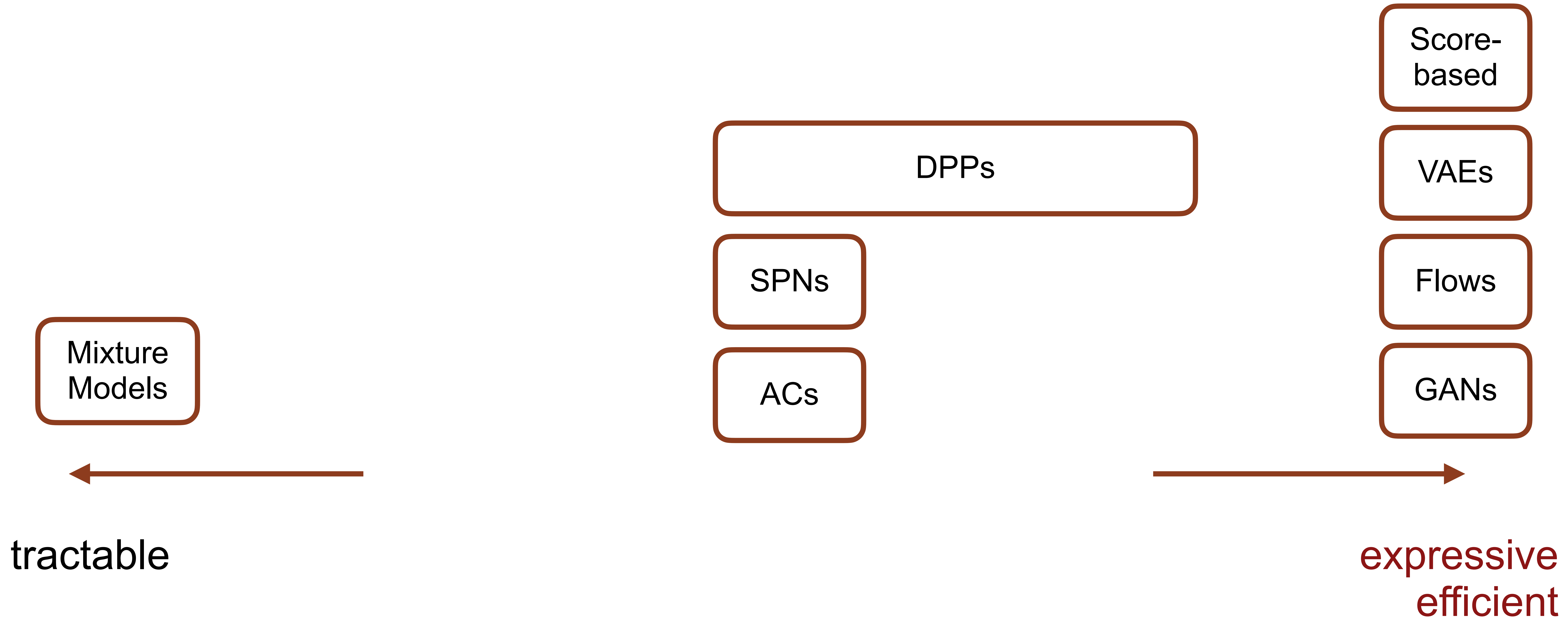
ACs

Mixture Models

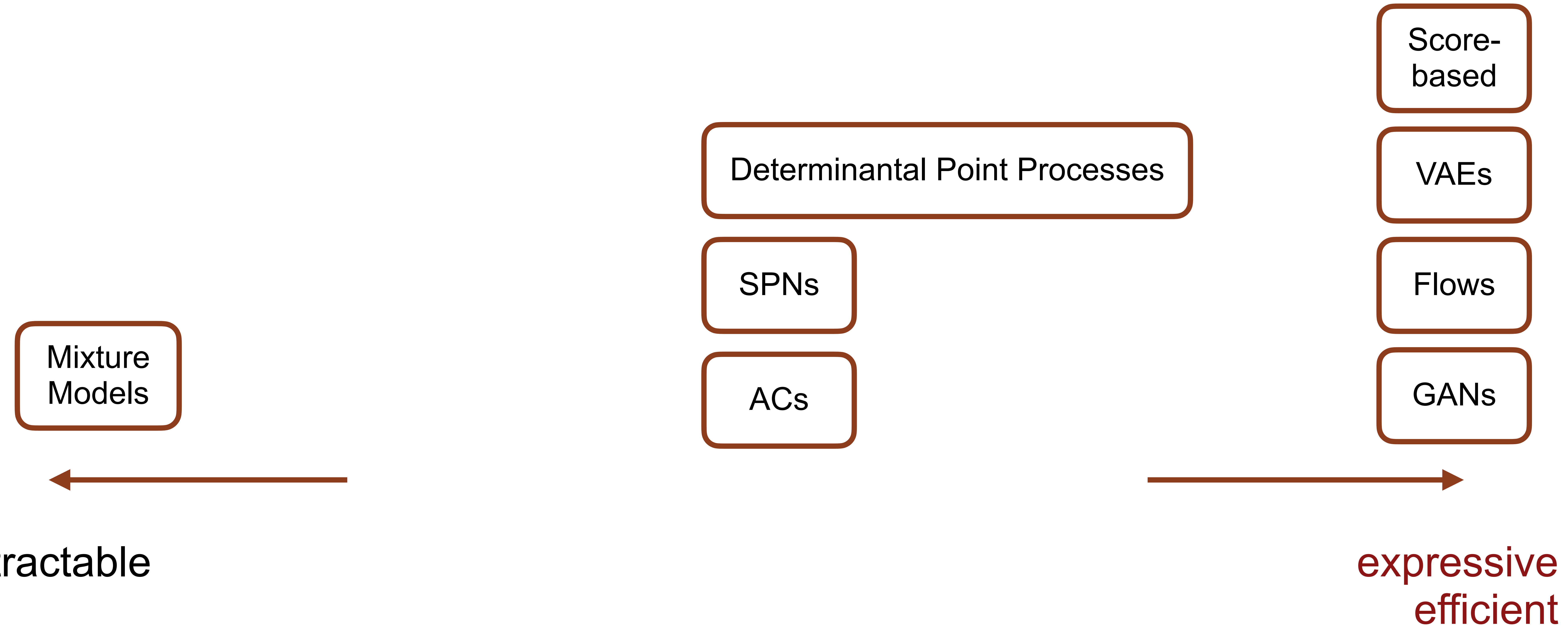
tractable

expressive
efficient

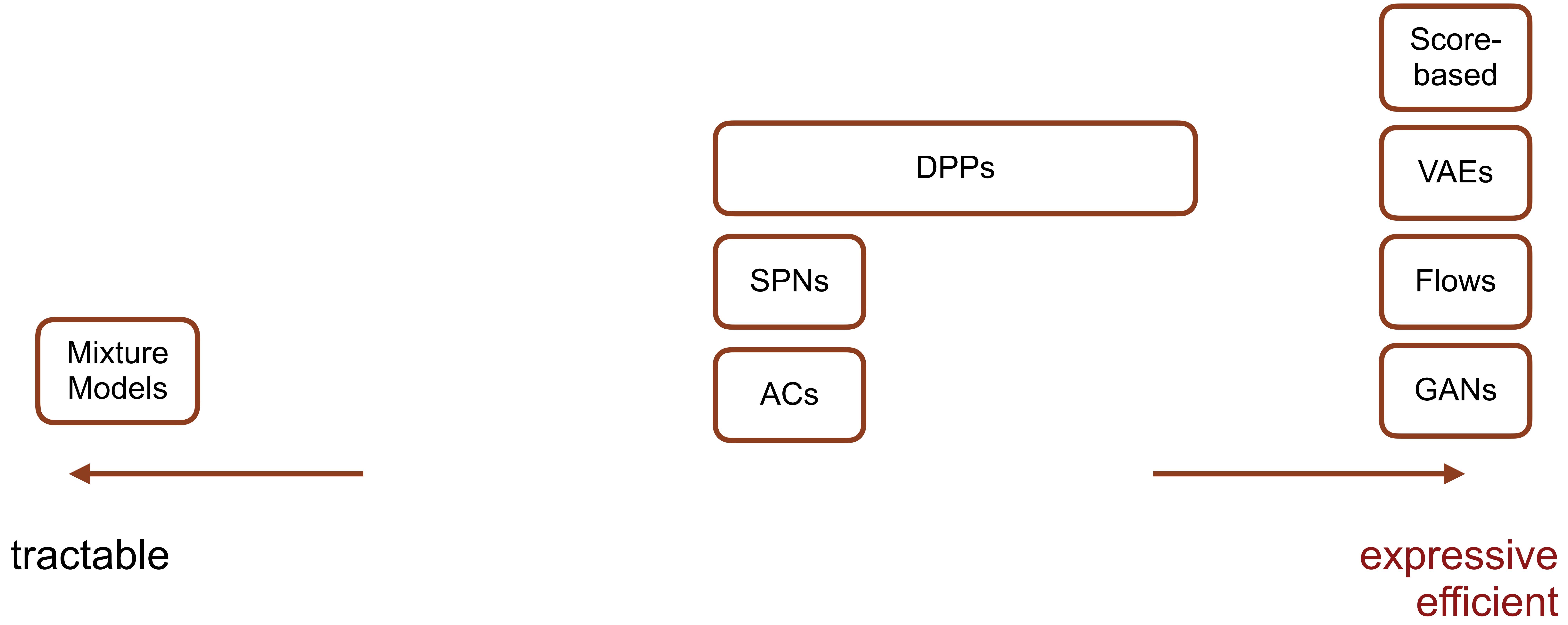
Modeling Families



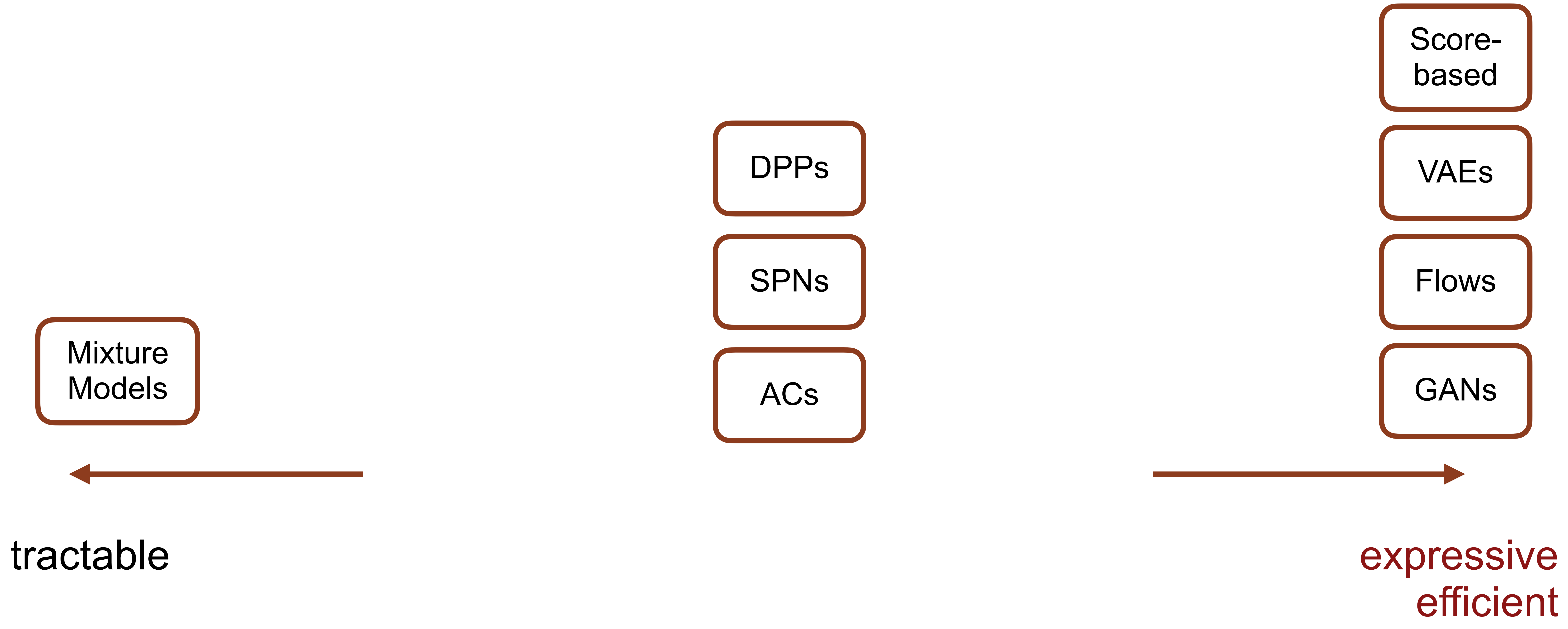
Modeling Families



Modeling Families



Modeling Families



Modeling Families

Score-based

VAEs

Flows

GANs

DPPs

SPNs

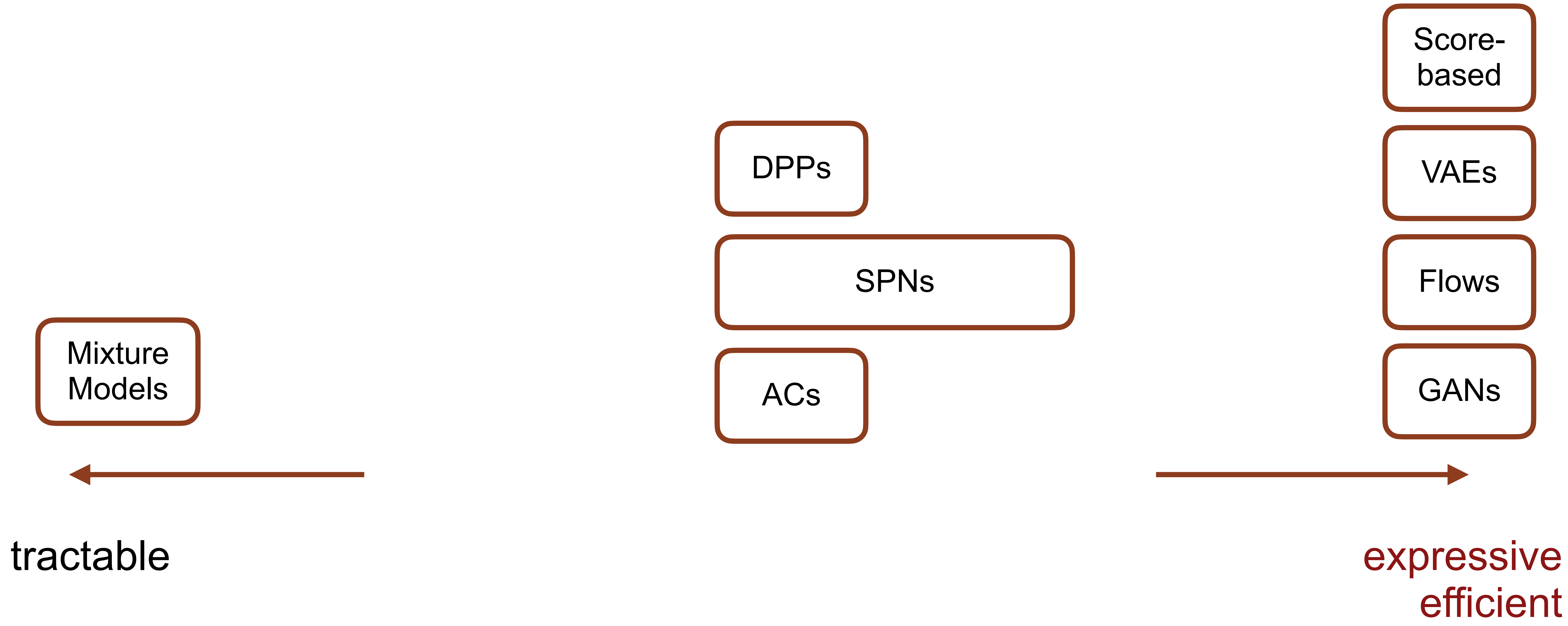
ACs

Mixture Models

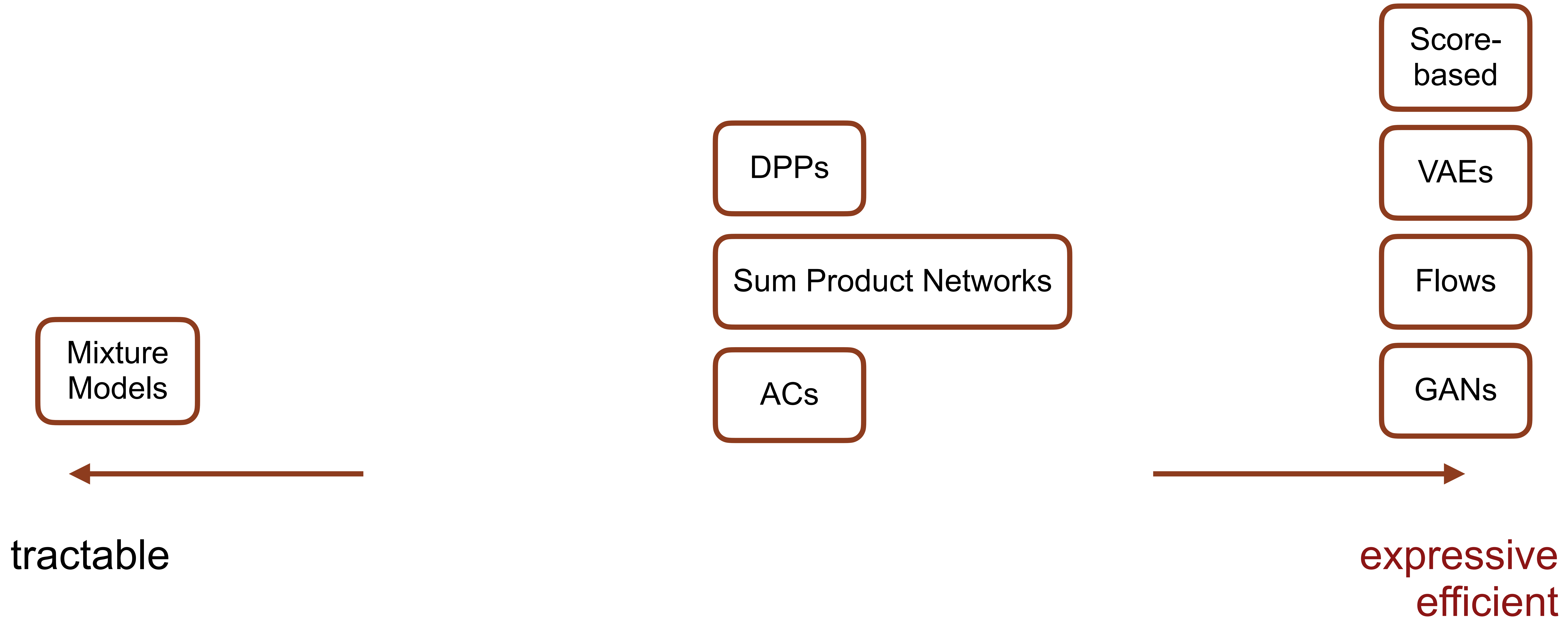
tractable

expressive
efficient

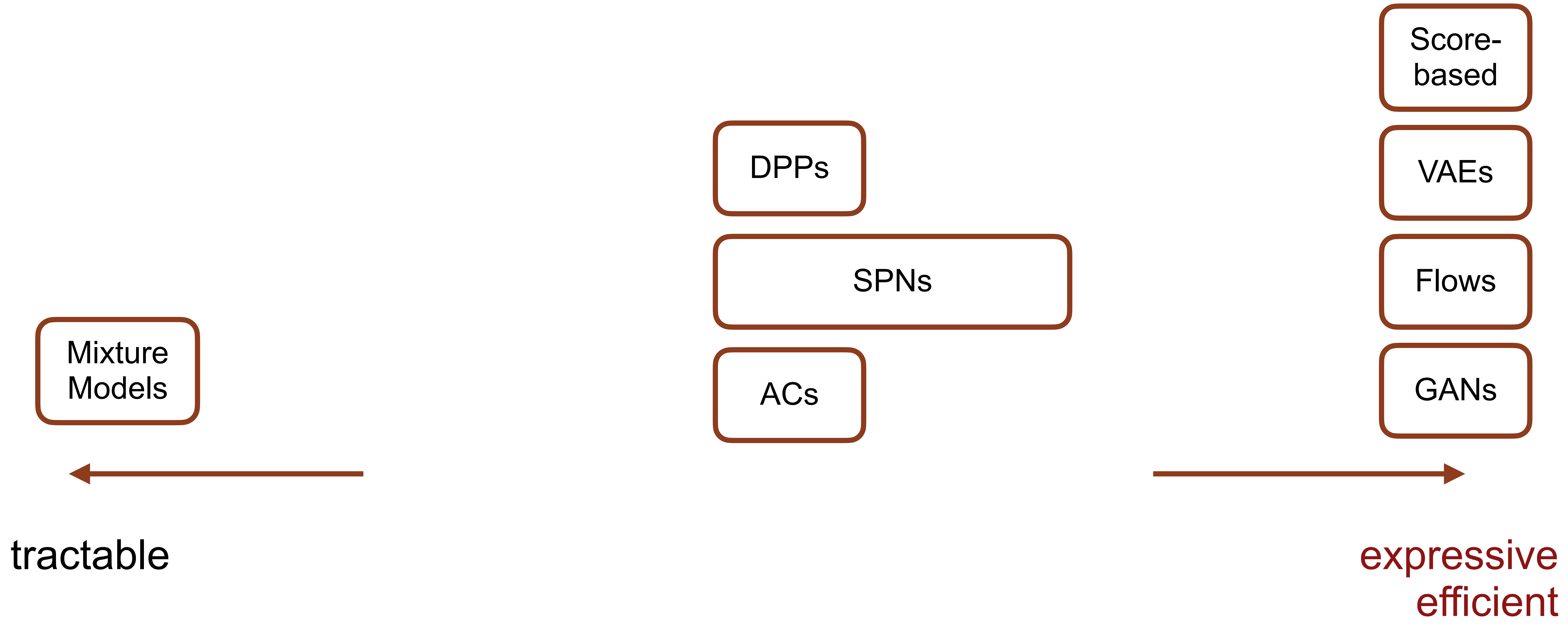
Modeling Families



Modeling Families



Modeling Families



Modeling Families

Mixture Models

DPPs

SPNs

ACs

Score-based

VAEs

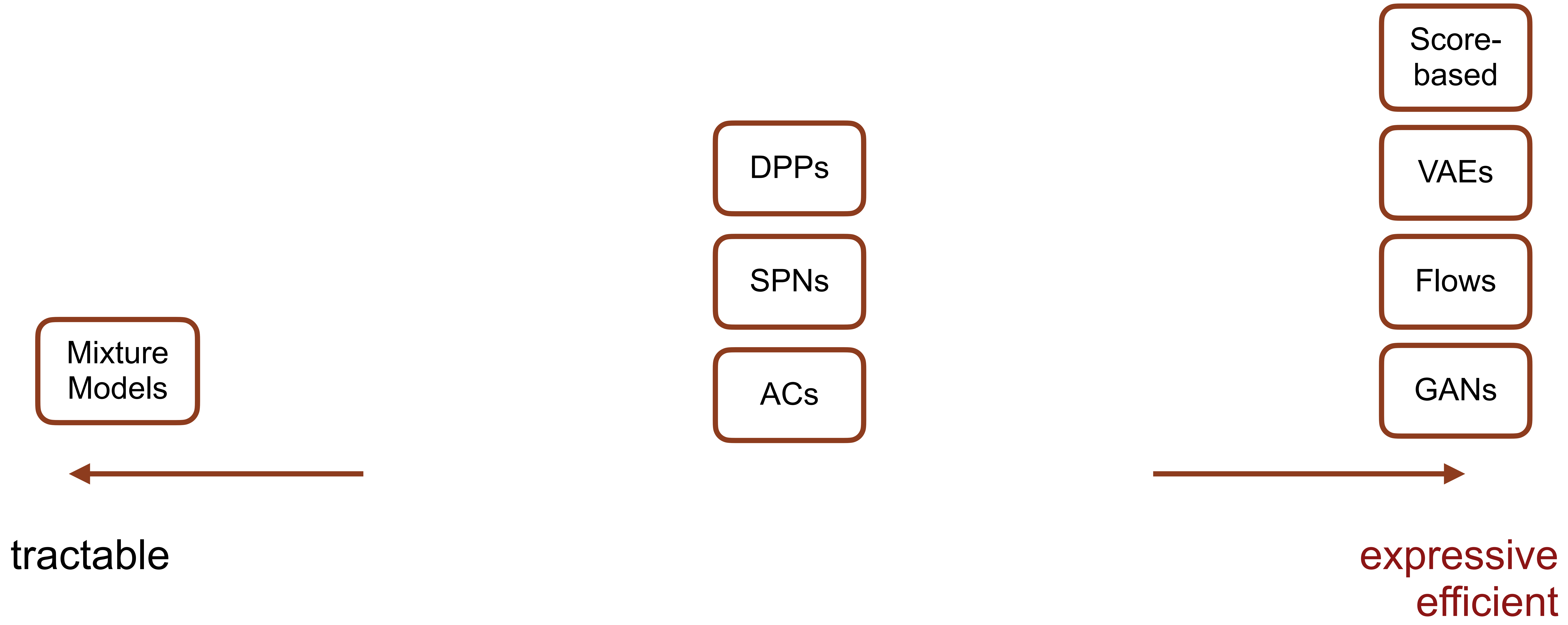
Flows

GANs

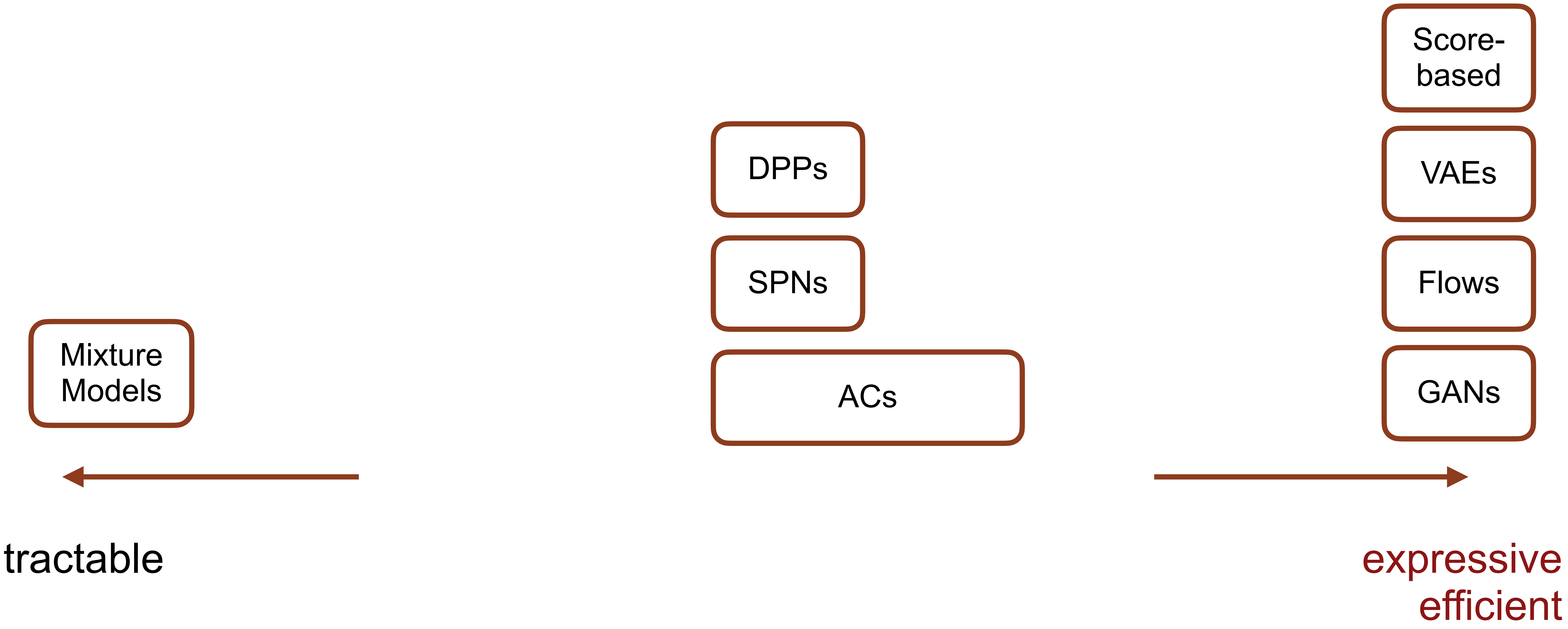
tractable

expressive
efficient

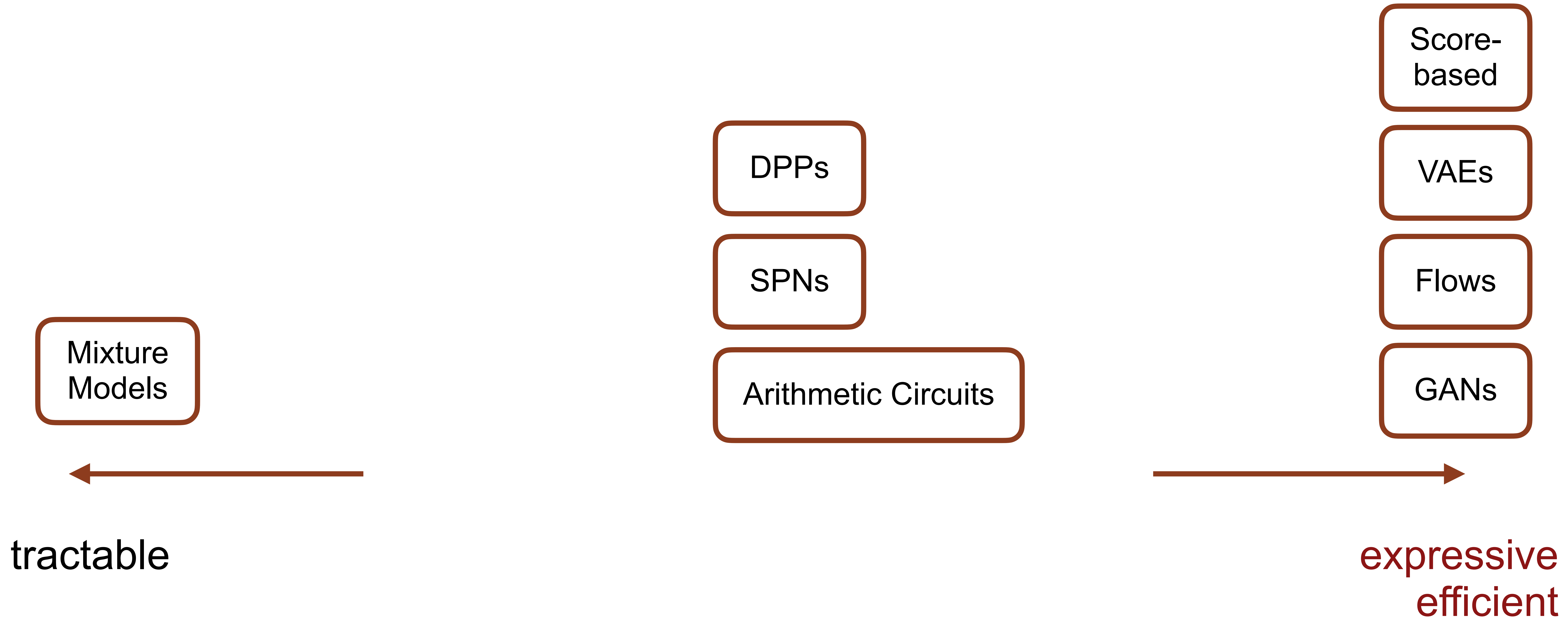
Modeling Families



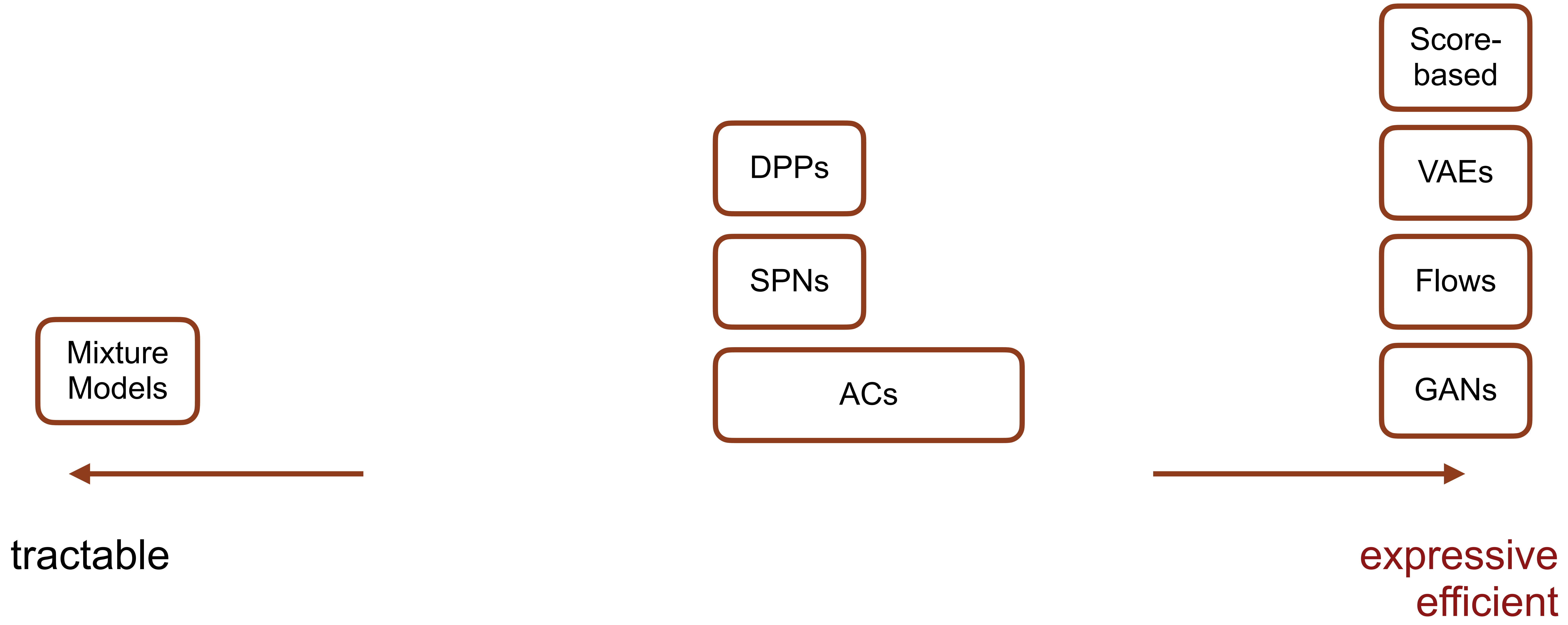
Modeling Families



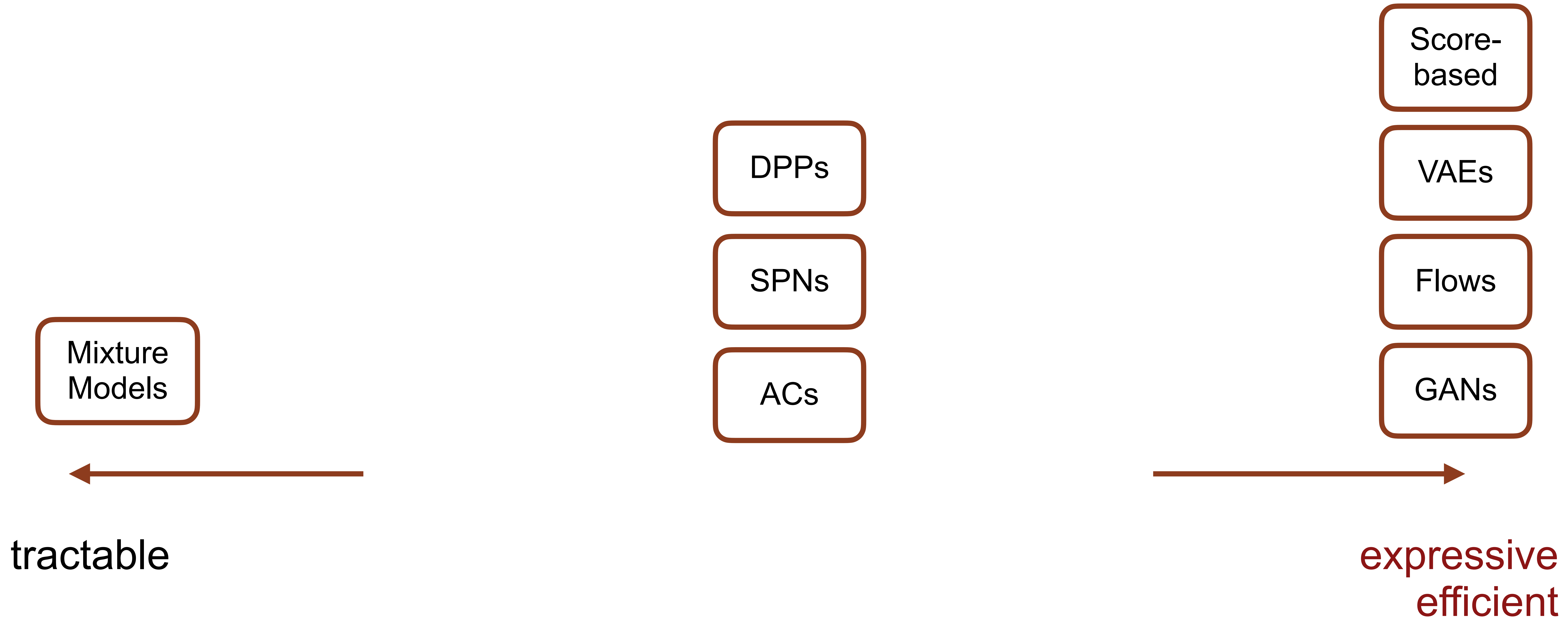
Modeling Families



Modeling Families



Modeling Families



Modeling Families

Sum Product Networks

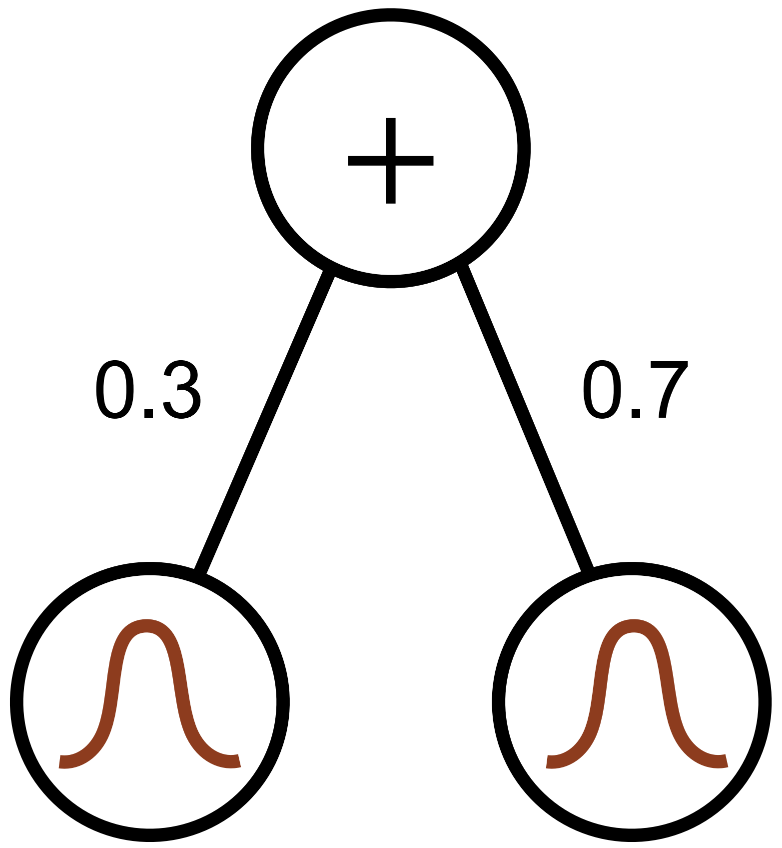
base distributions

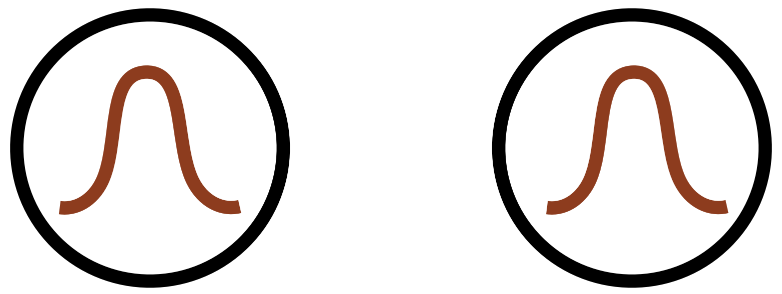
Sum Product Networks

sum nodes

mixture

base
distributions

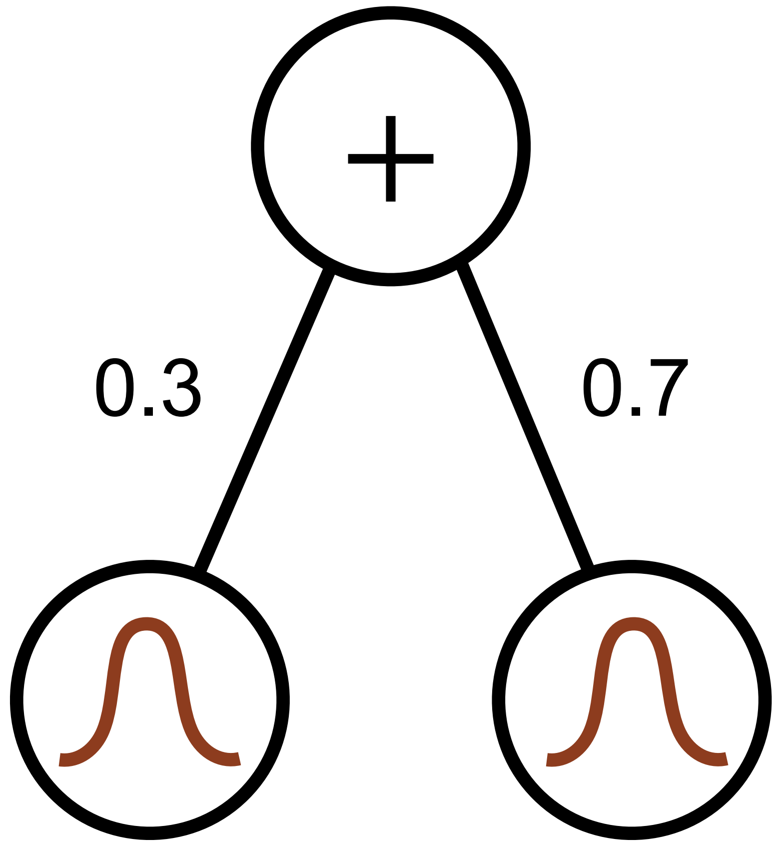




Sum Product Networks

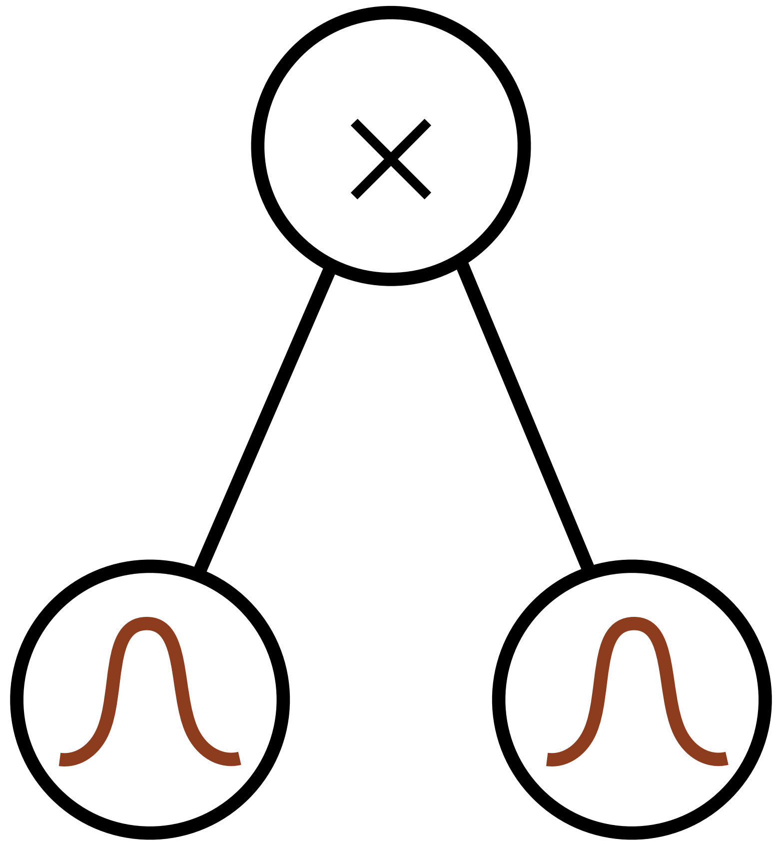
sum nodes

mixture



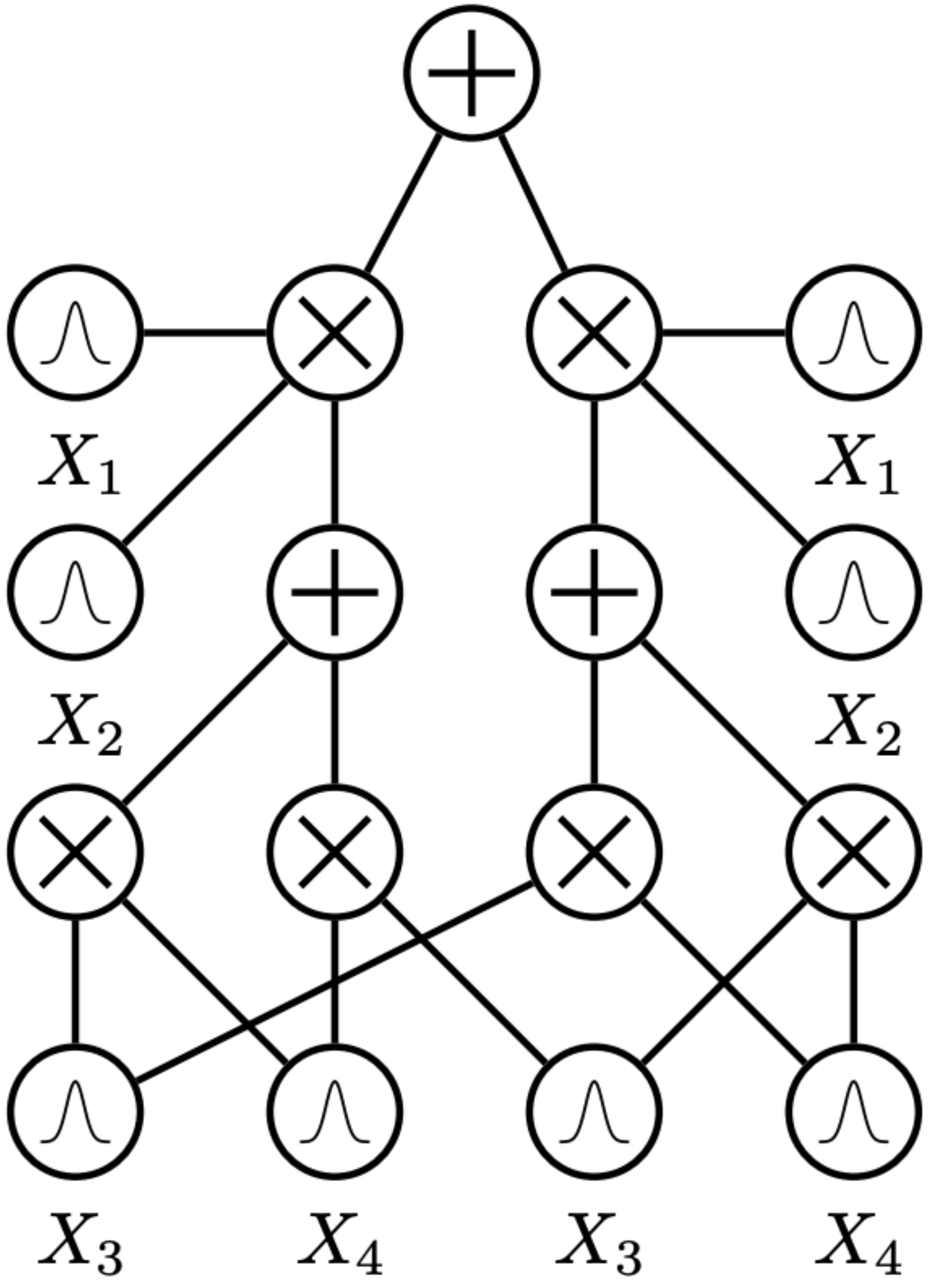
product nodes

factorization



base distributions

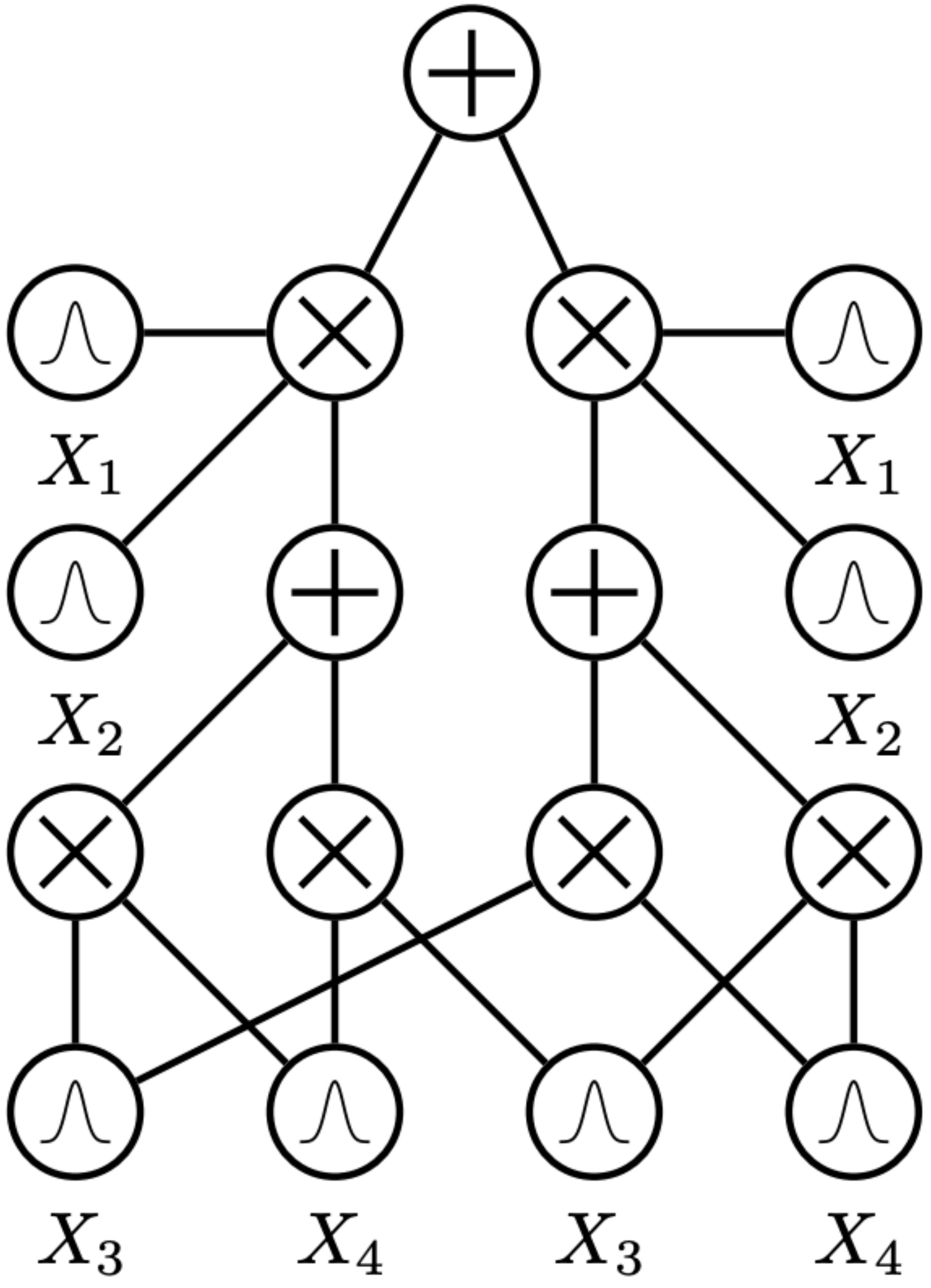
Sum Product Networks



Sum Product Networks

Feed-forward network

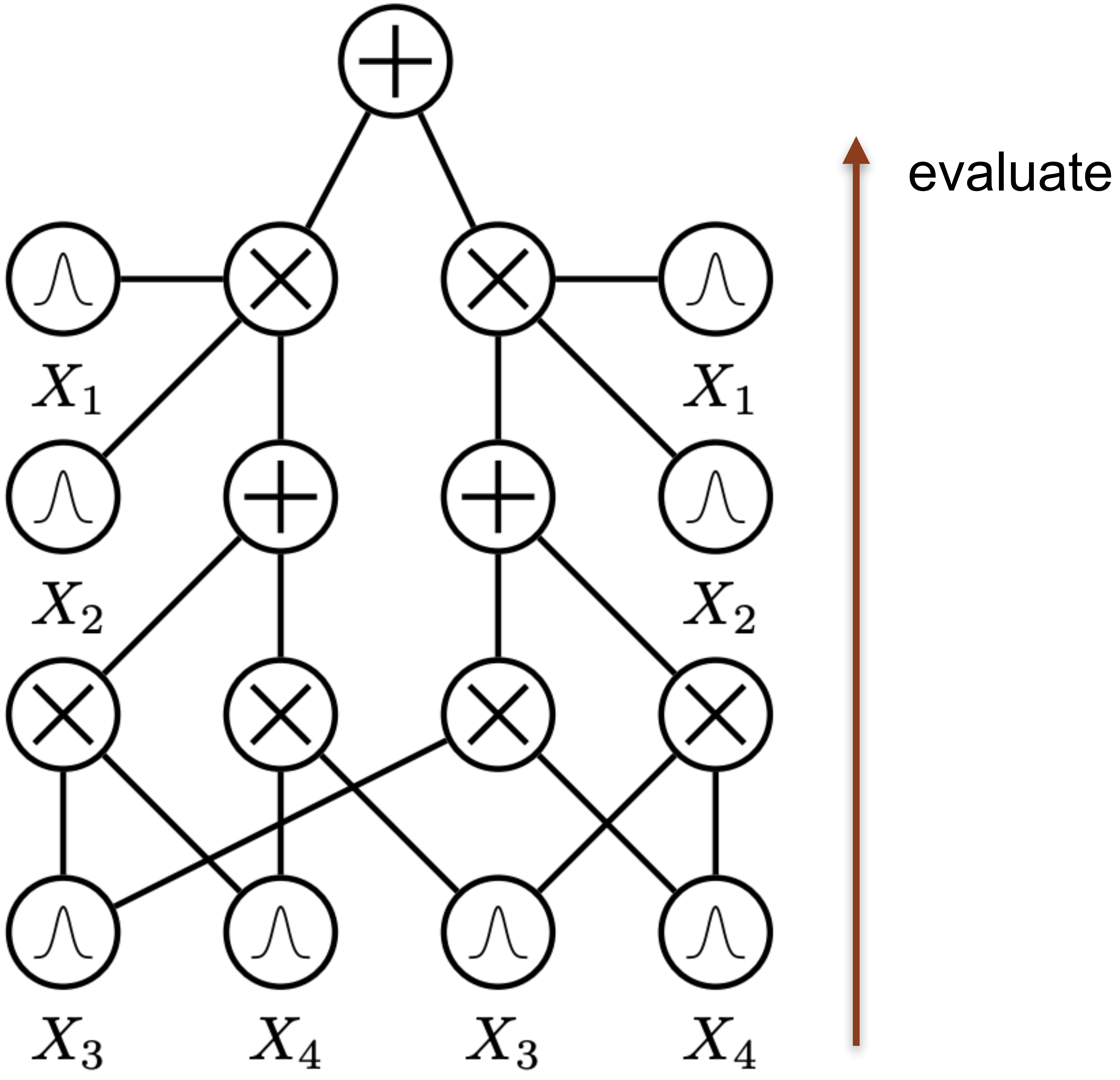
Defines a computation graph



Sum Product Networks

Feed-forward network

Defines a computation graph

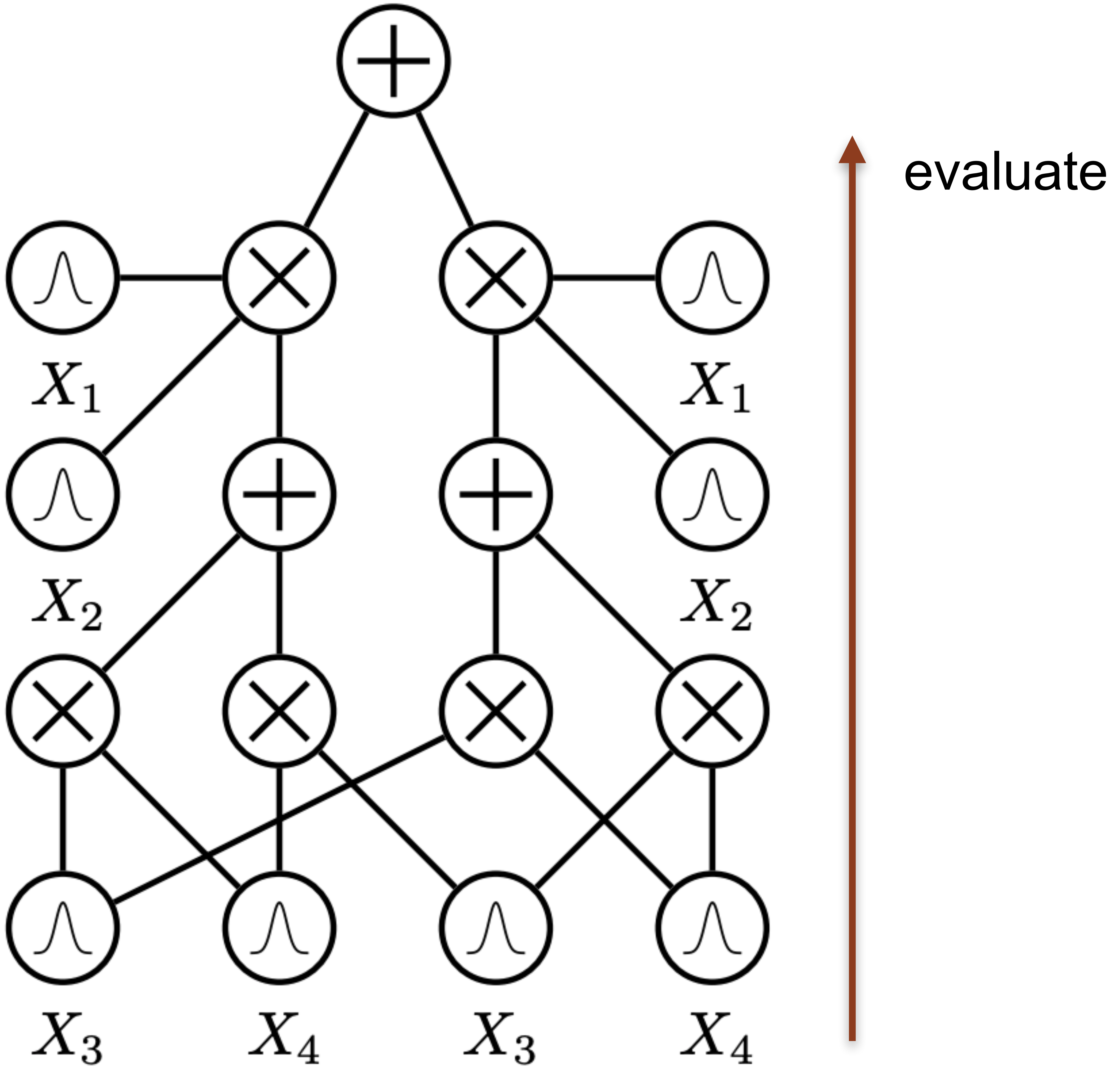


Sum Product Networks

Feed-forward network

Defines a computation graph

Train via gradient descent

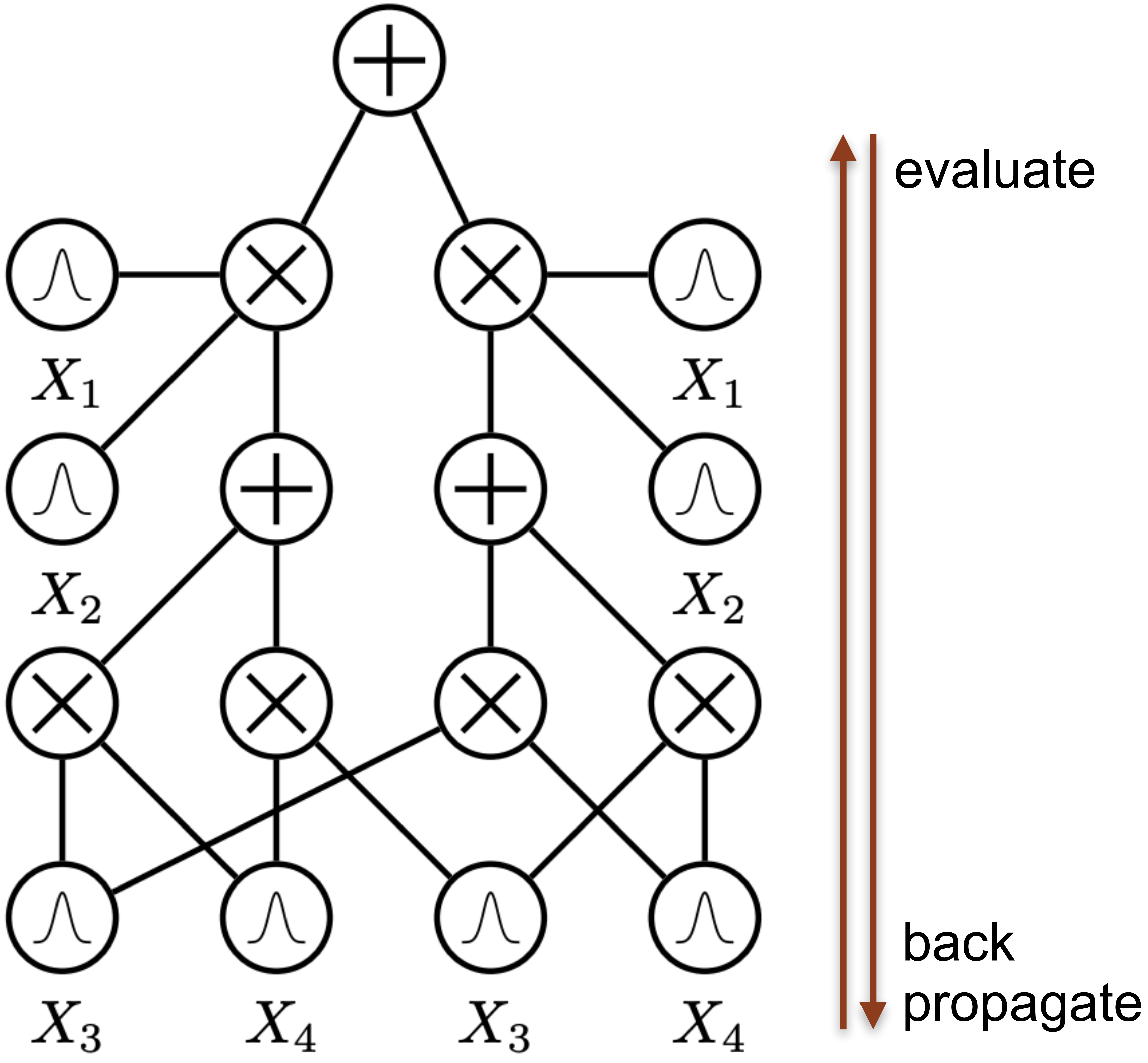


Sum Product Networks

Feed-forward network

Defines a computation graph

Train via gradient descent



Sum Product Networks

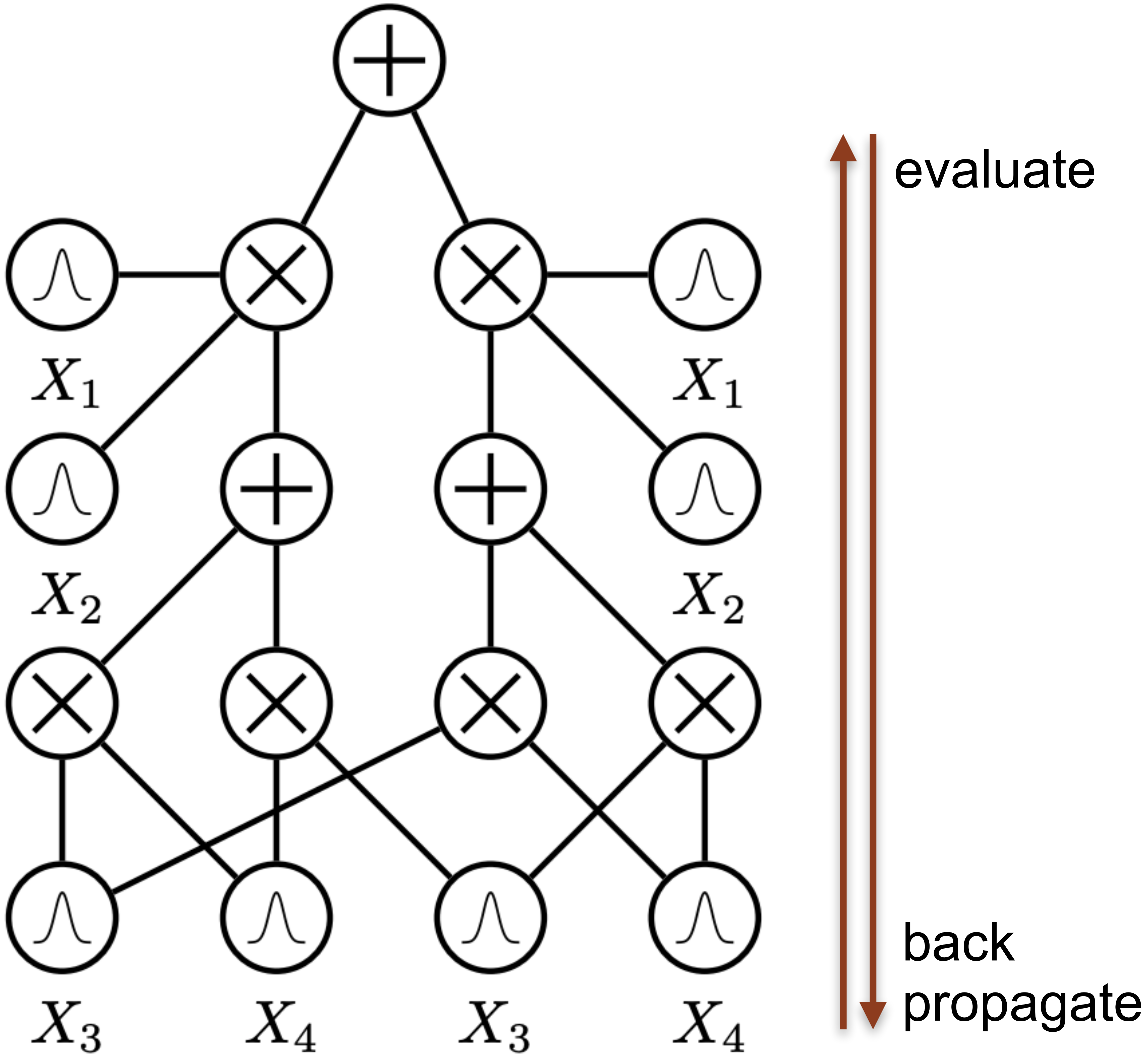
Feed-forward network

Defines a computation graph

Train via gradient descent

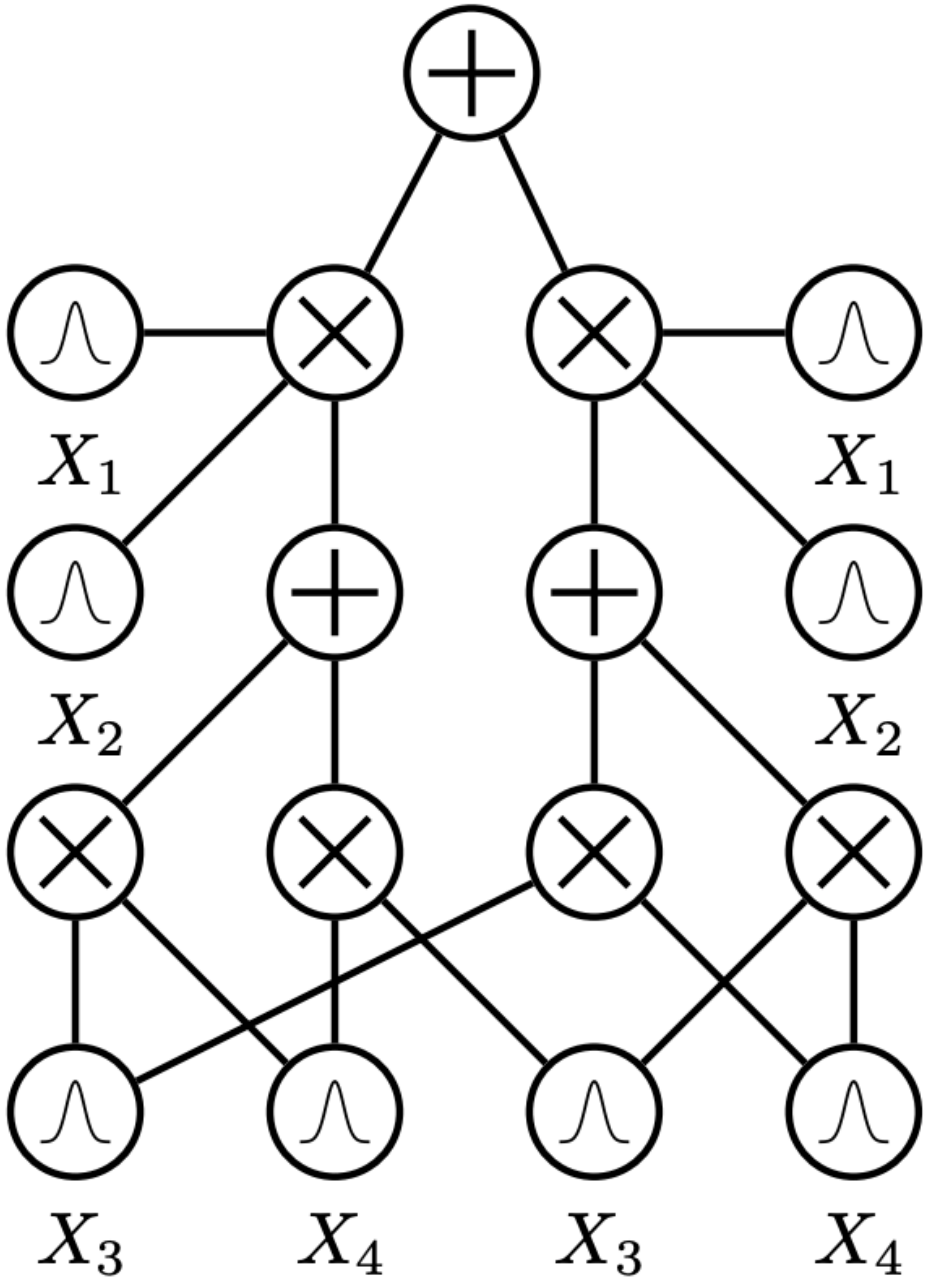
Multilinear polynomial
over base distributions

✓ tractable



Sum Product Networks

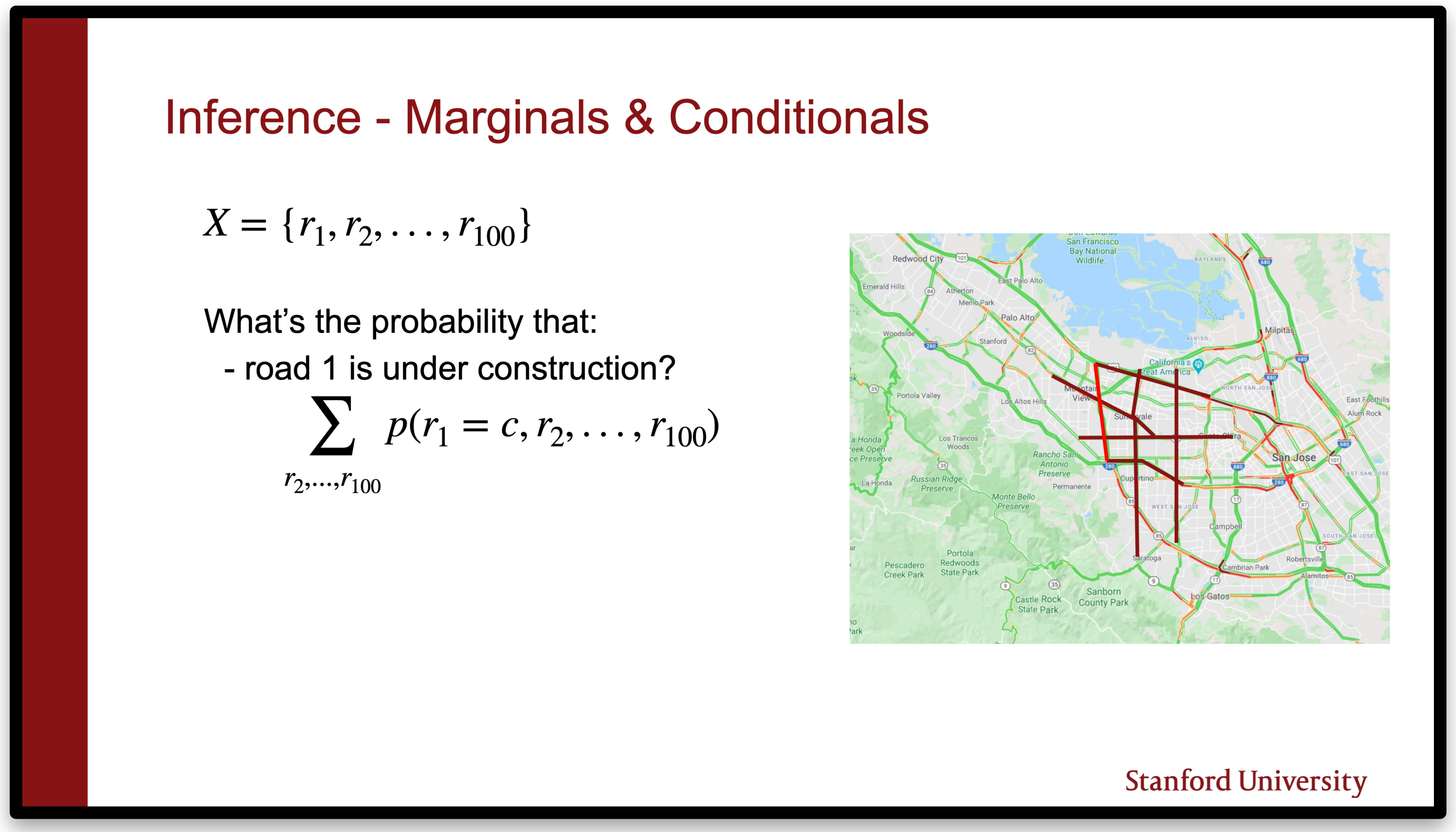
✓ Marginals in one forward pass

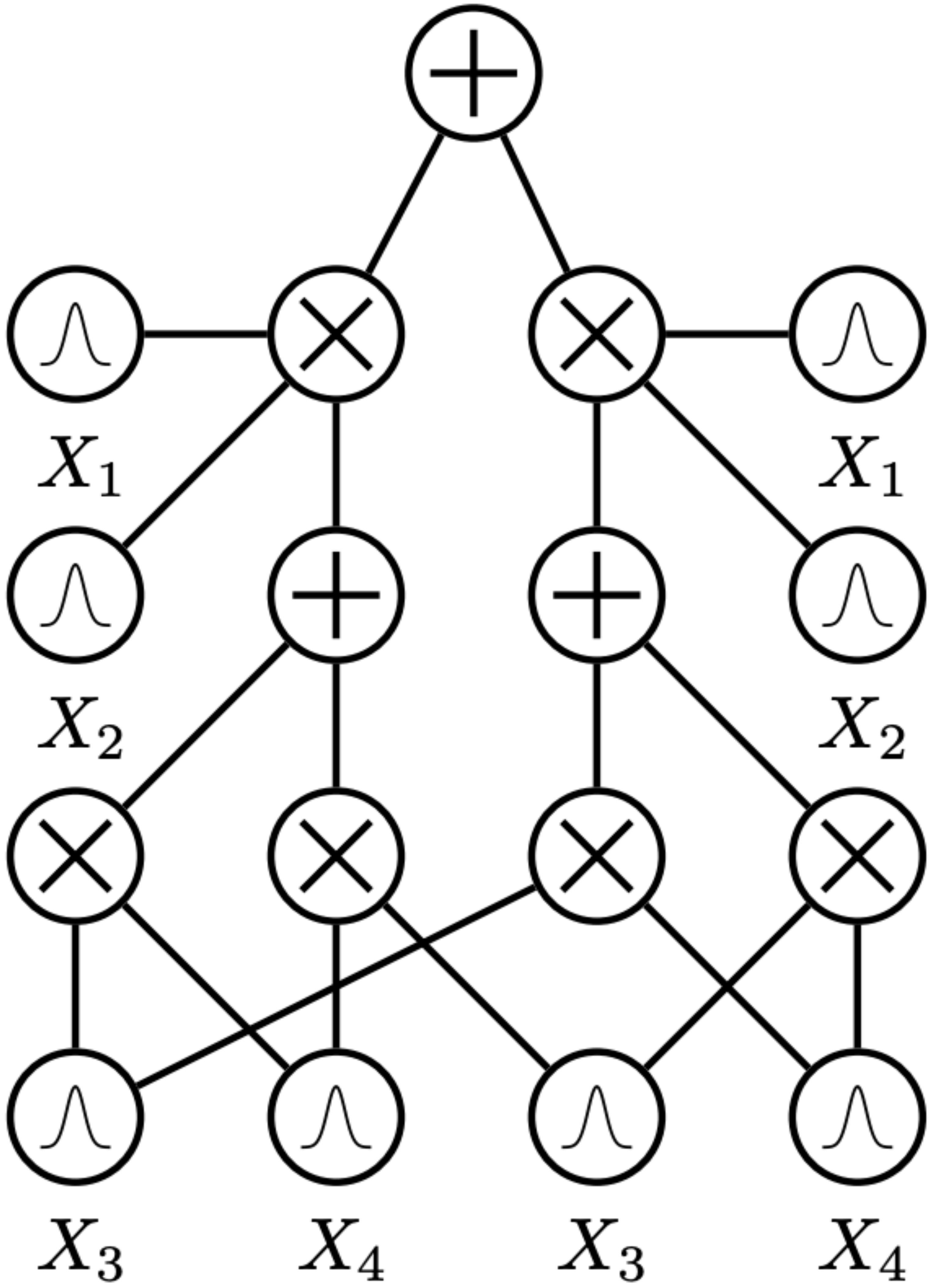


Sum Product Networks

✔ Marginals in one forward pass

Recall…

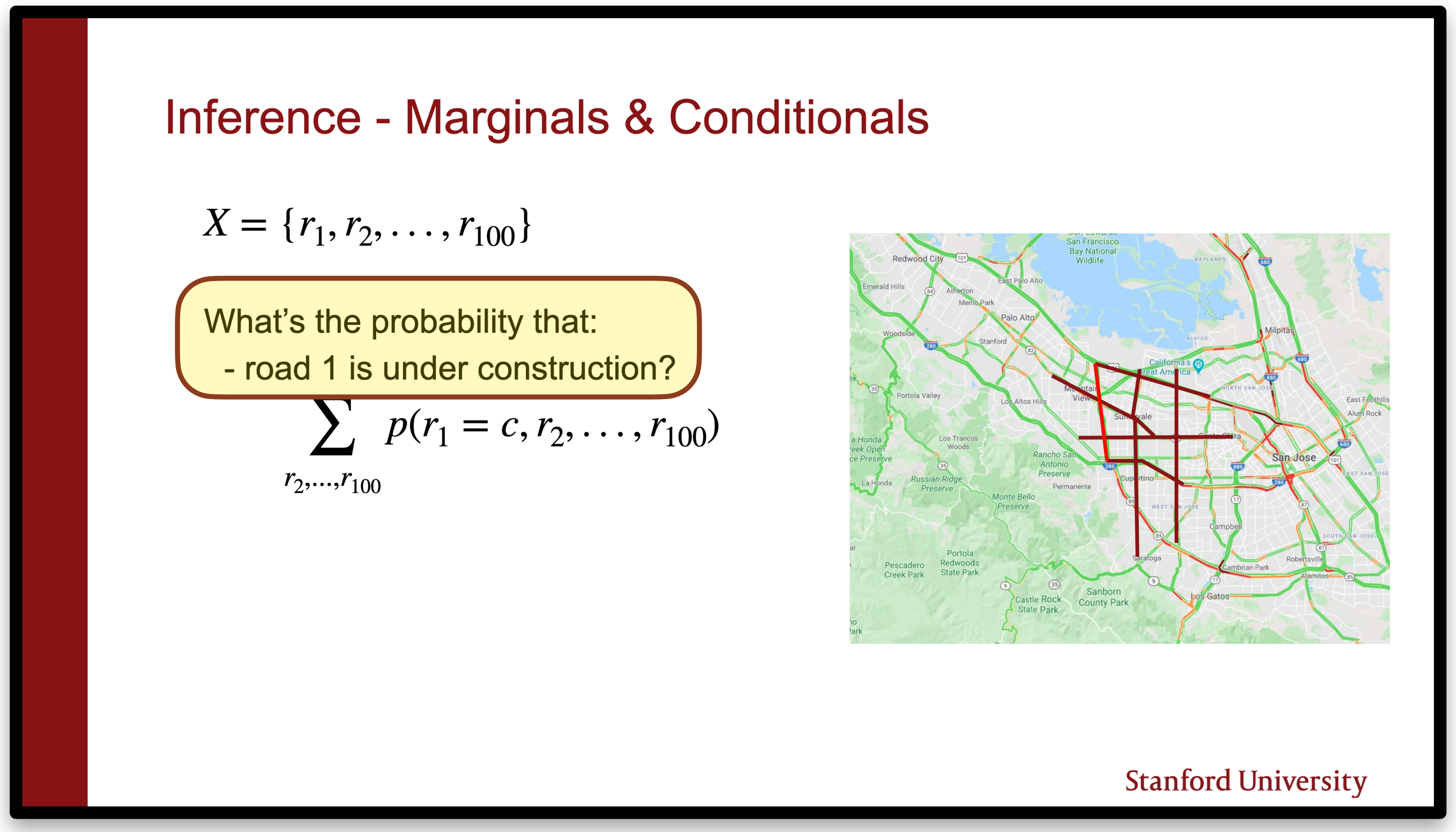


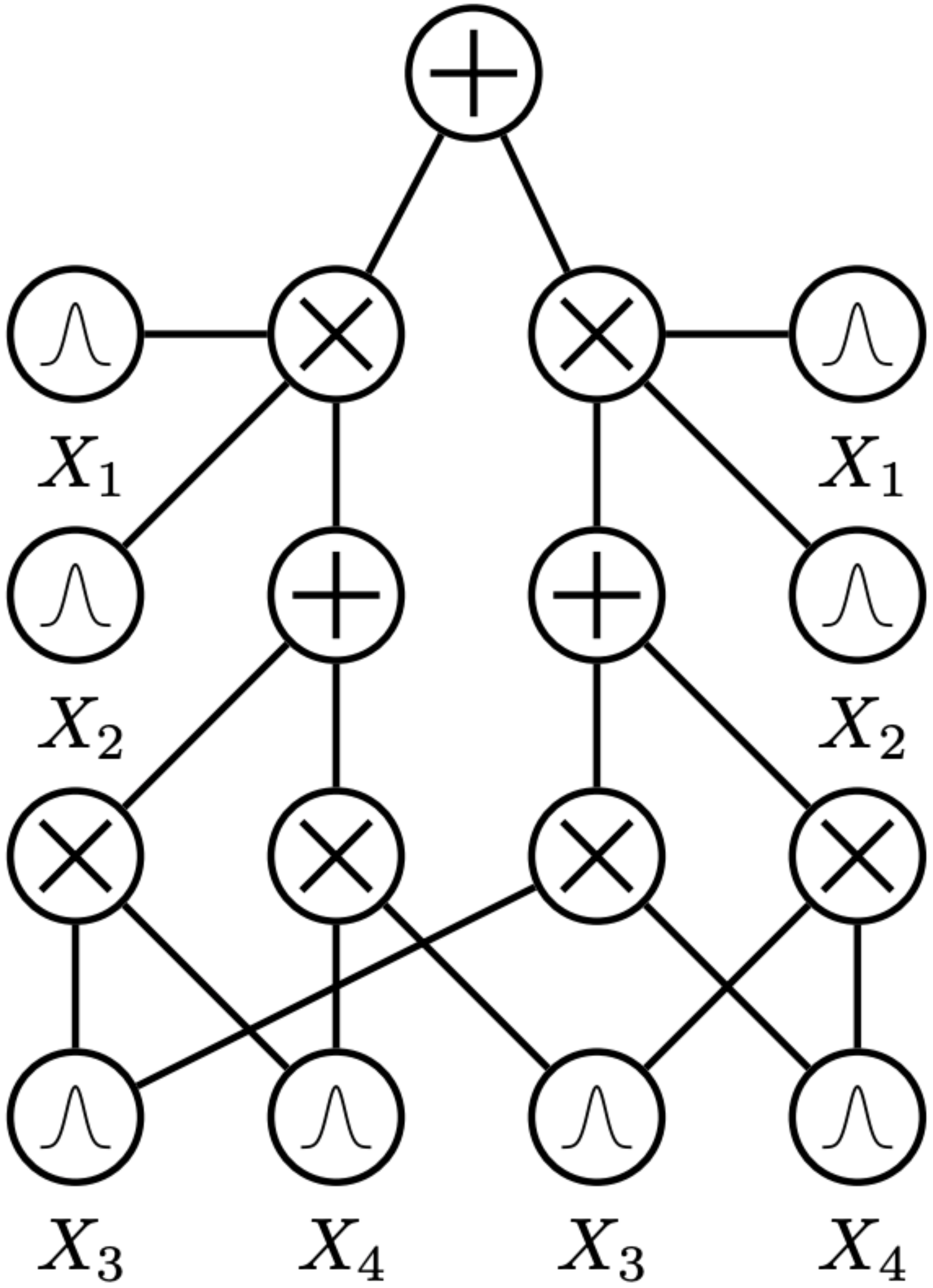


Sum Product Networks

✓ Marginals in one forward pass

Recall…

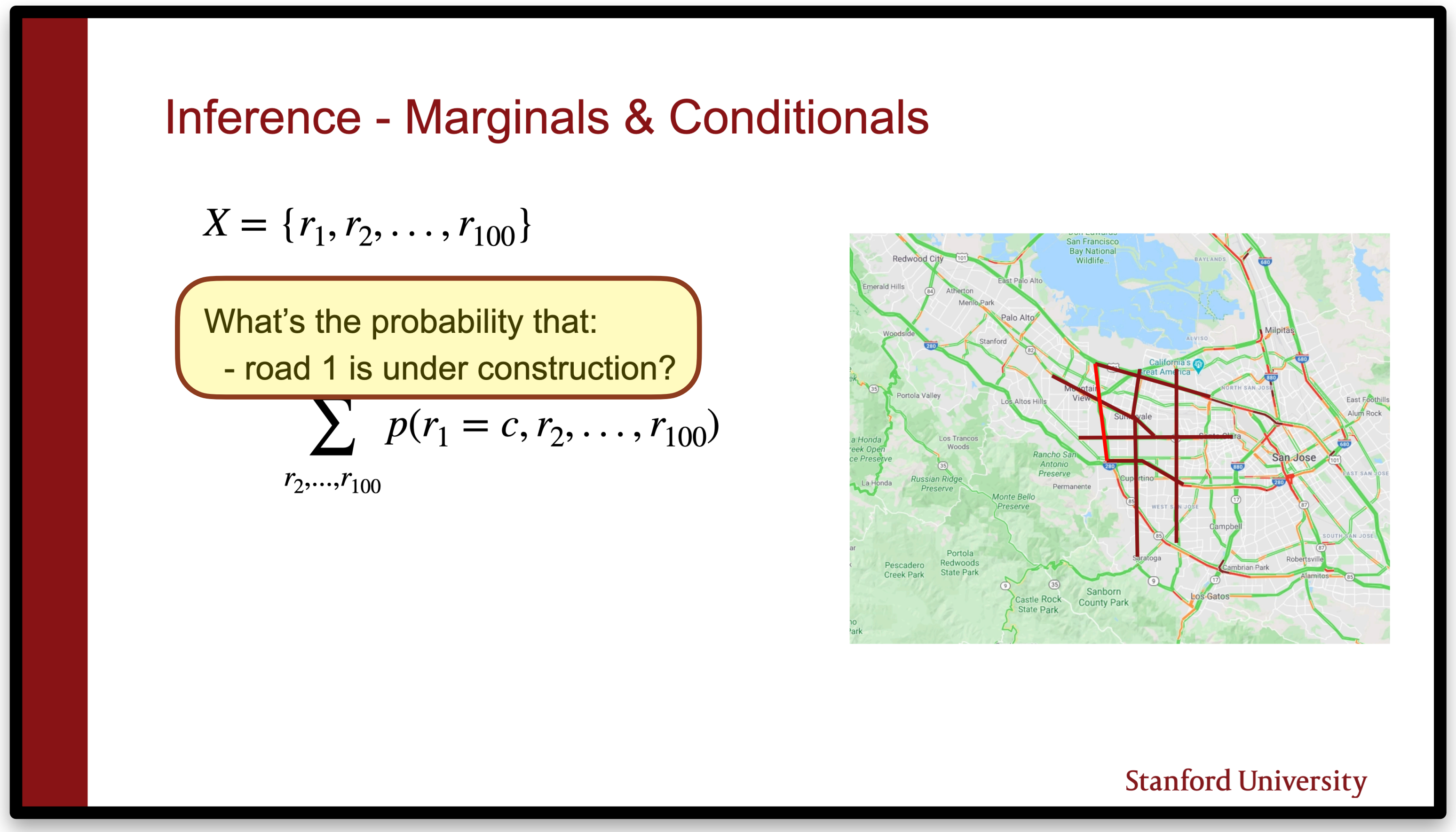


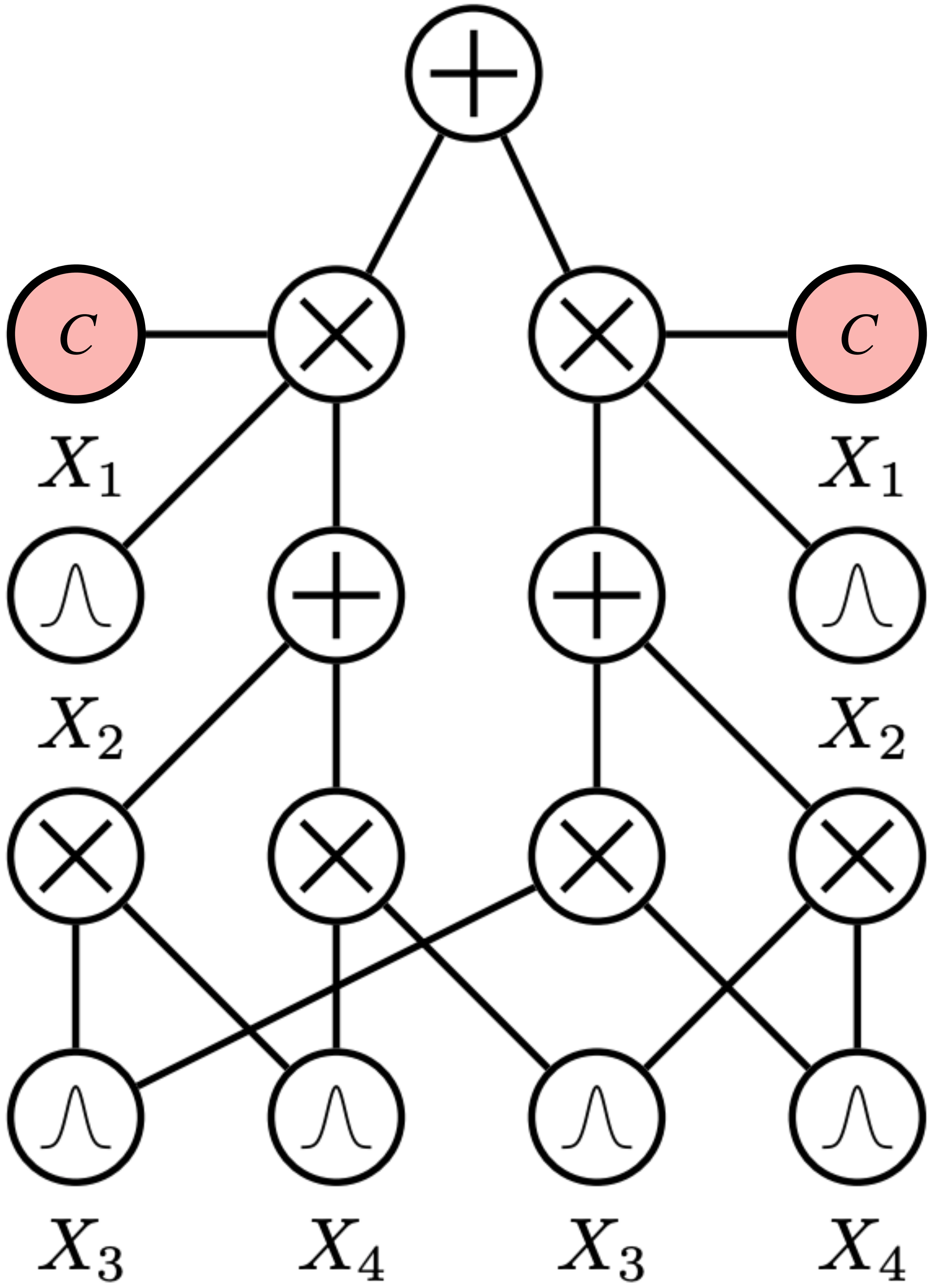


Sum Product Networks

✓ Marginals in one forward pass

Recall…

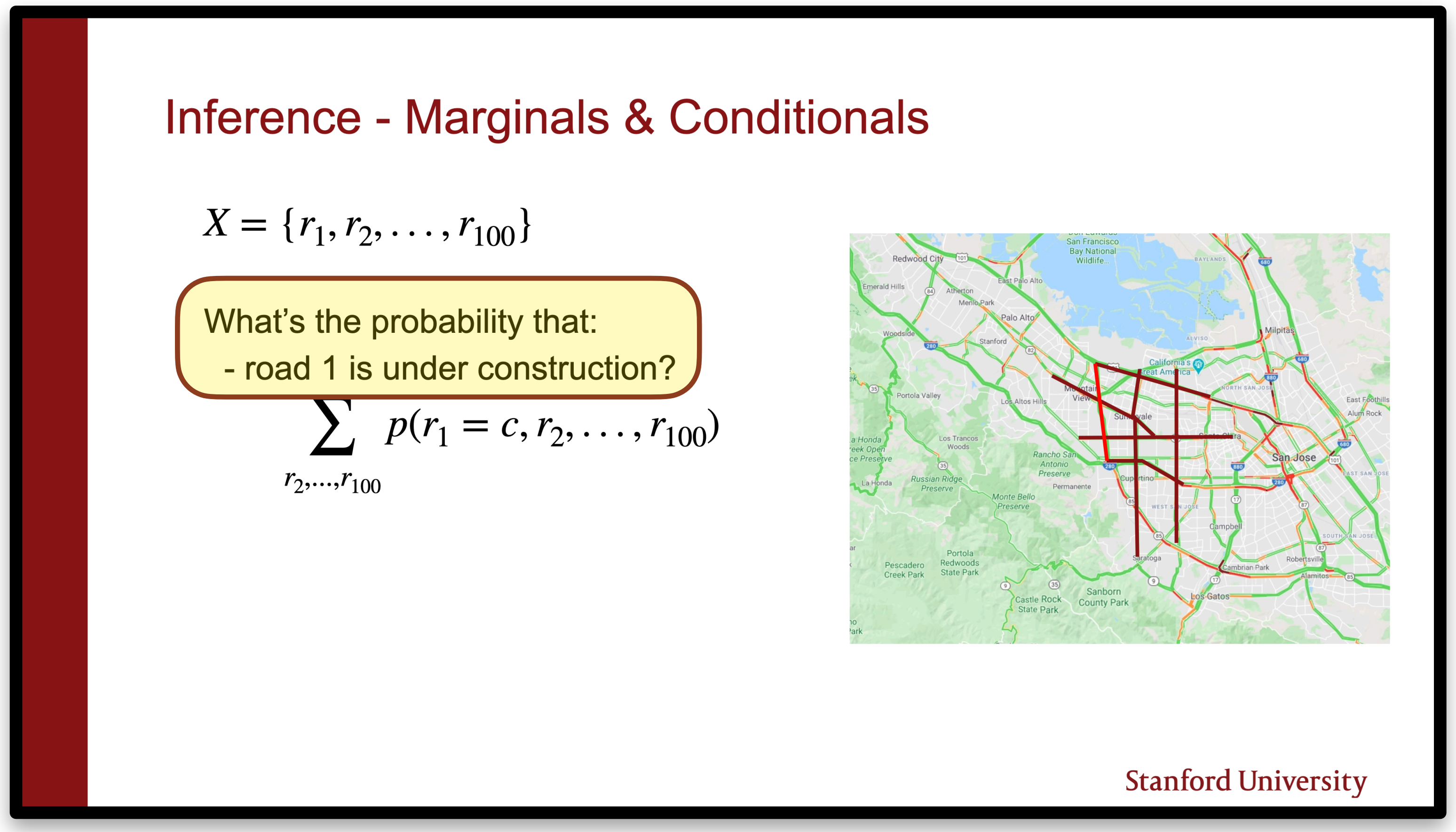


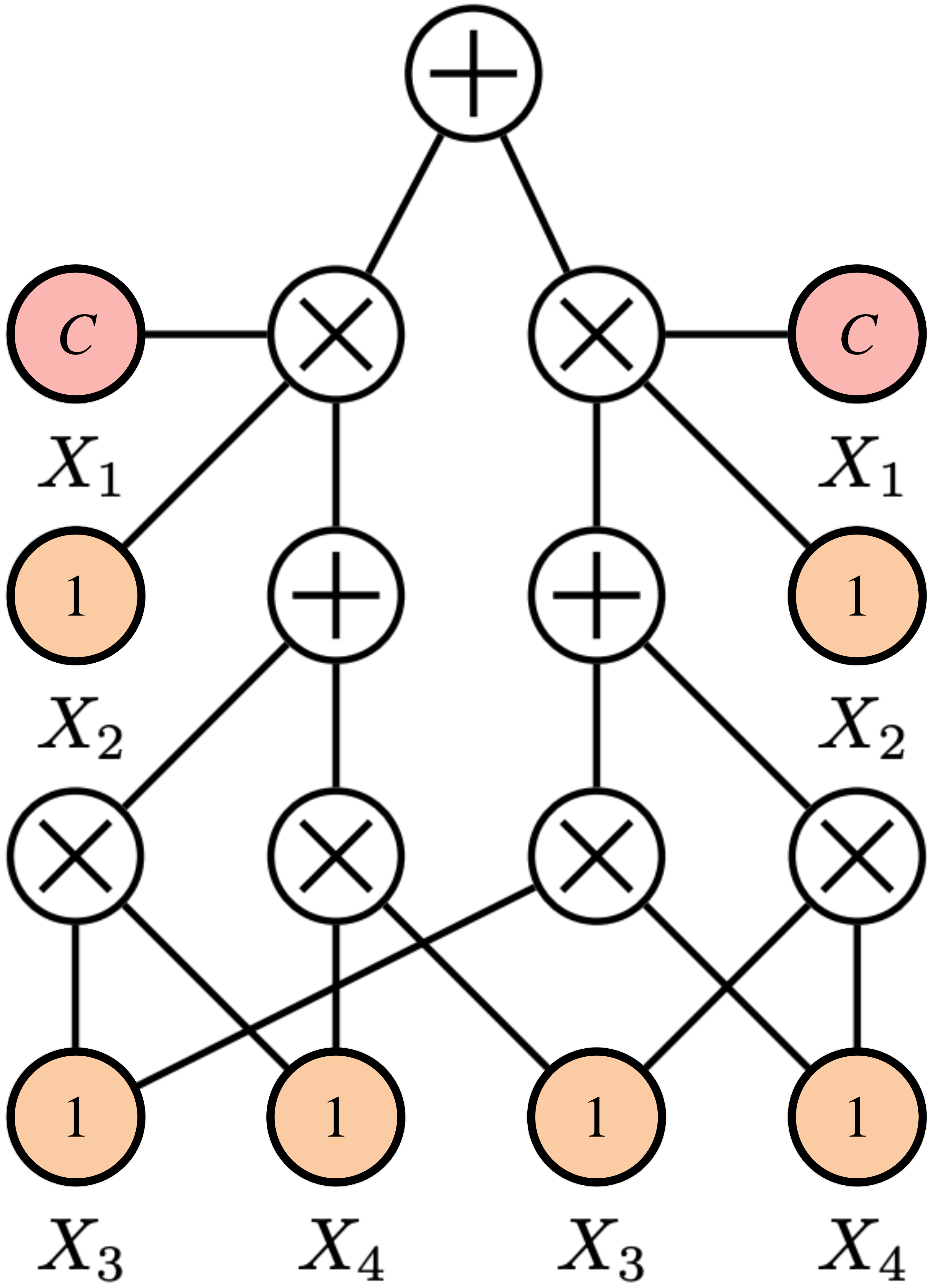


Sum Product Networks

✓ Marginals in one forward pass

Recall…

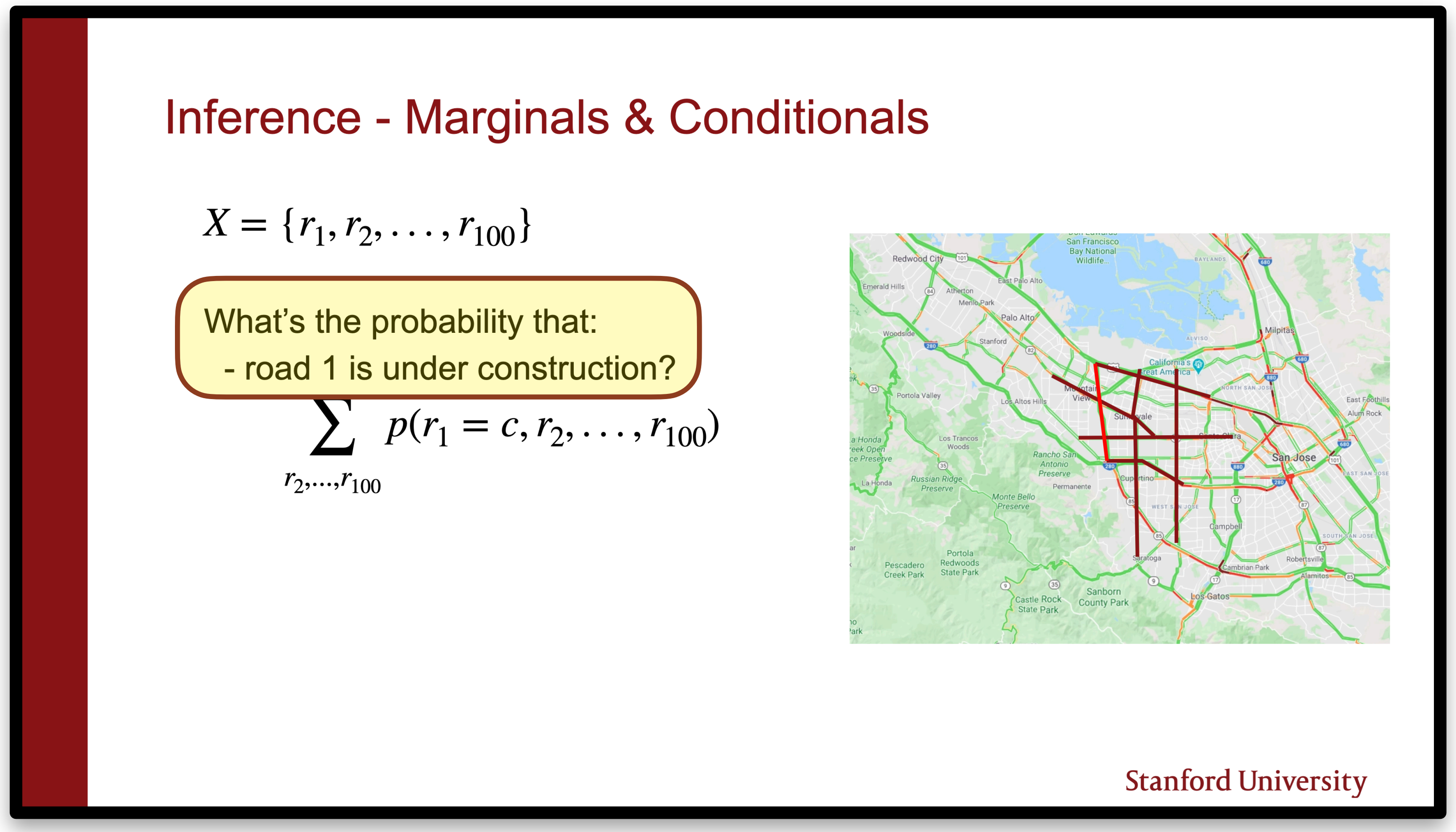


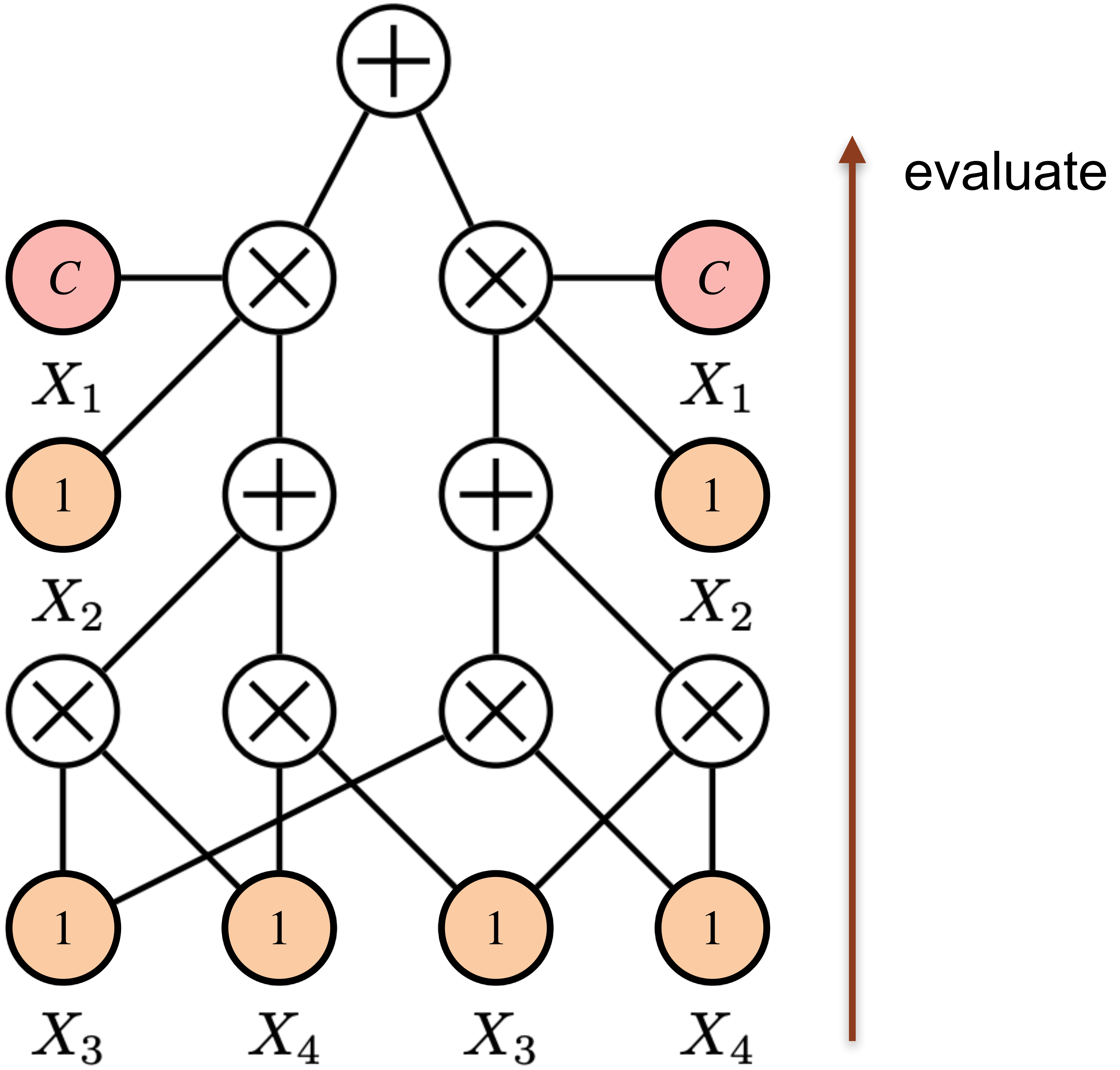


Sum Product Networks

✓ Marginals in one forward pass

Recall…





SPN Architecture

Structure Learning — expensive

SPN Architecture

Structure Learning — expensive

Prescribed Structure

RAT-SPNs (UAI'19)

EiNETs (ICML'20)

- 9.4M parameters

SPN Architecture

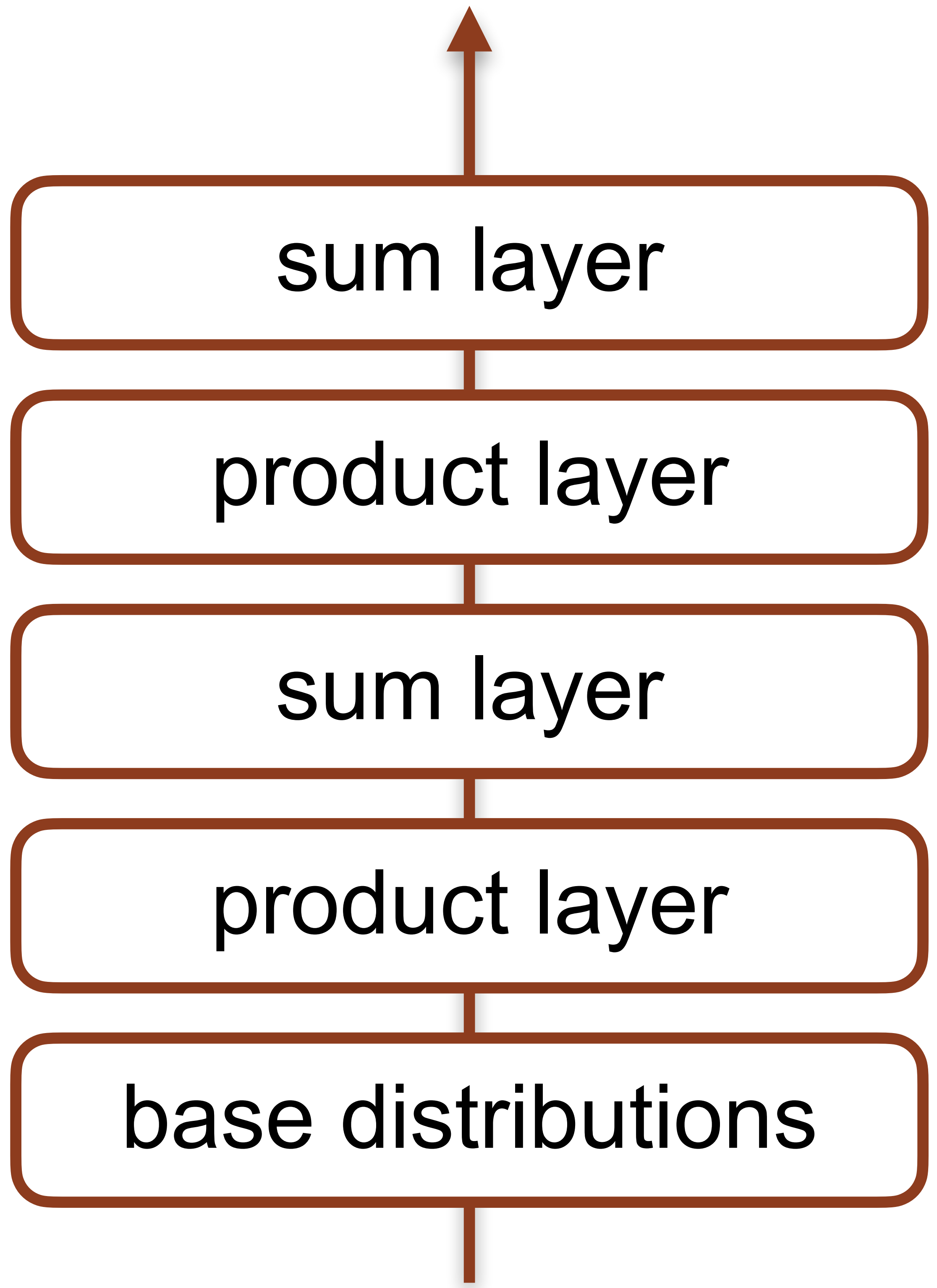
Structure Learning — expensive

Prescribed Structure

RAT-SPNs (UAI'19)

EiNETs (ICML'20)

- 9.4M parameters



Modeling Families

Score-based

DPPs

VAEs

SPNs

Flows

Mixture Models

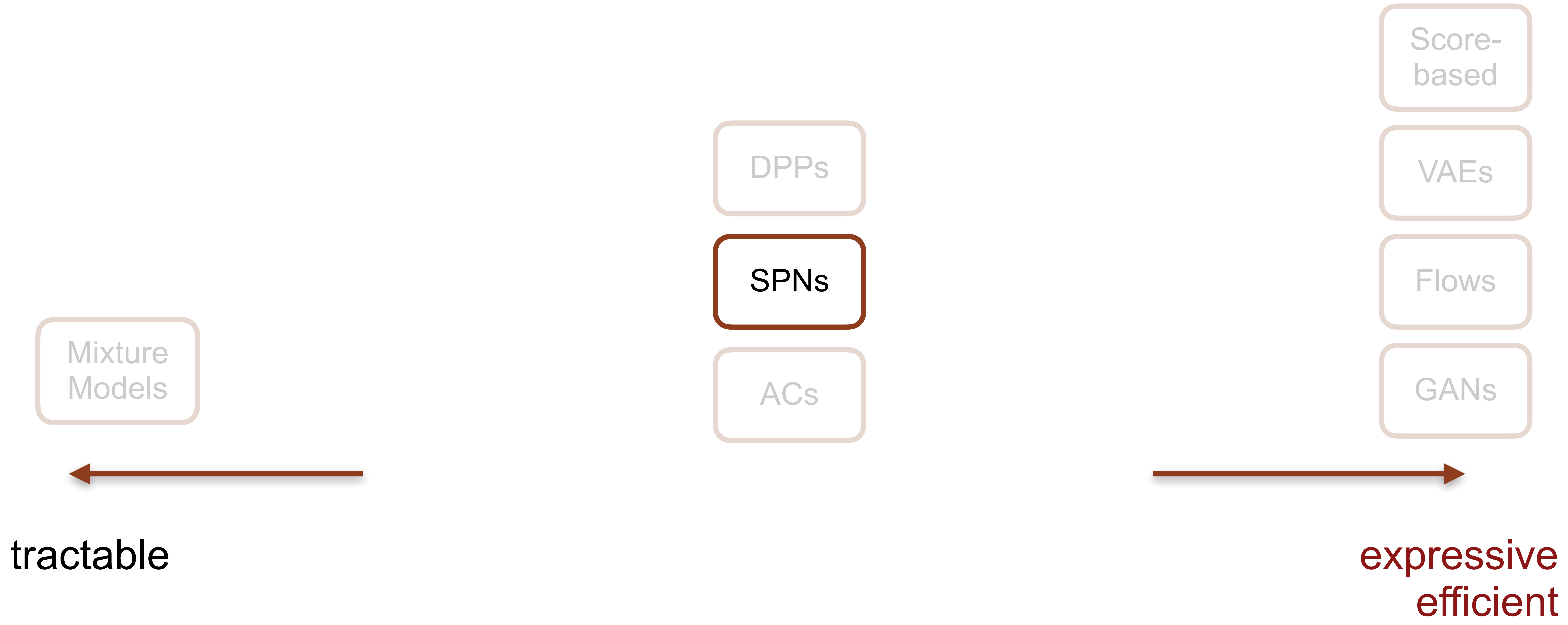
ACs

GANs

tractable

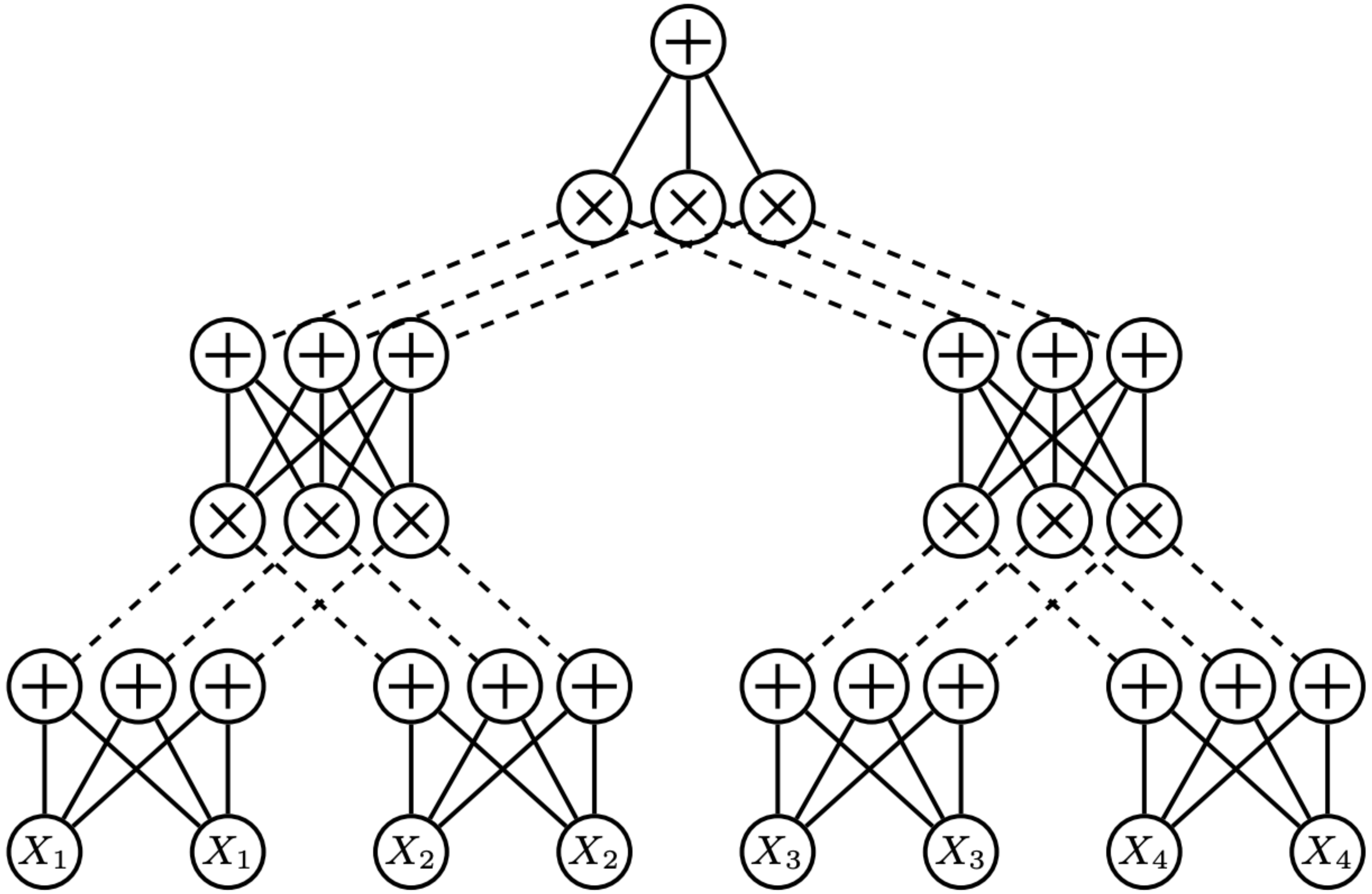
expressive
efficient

Modeling Families

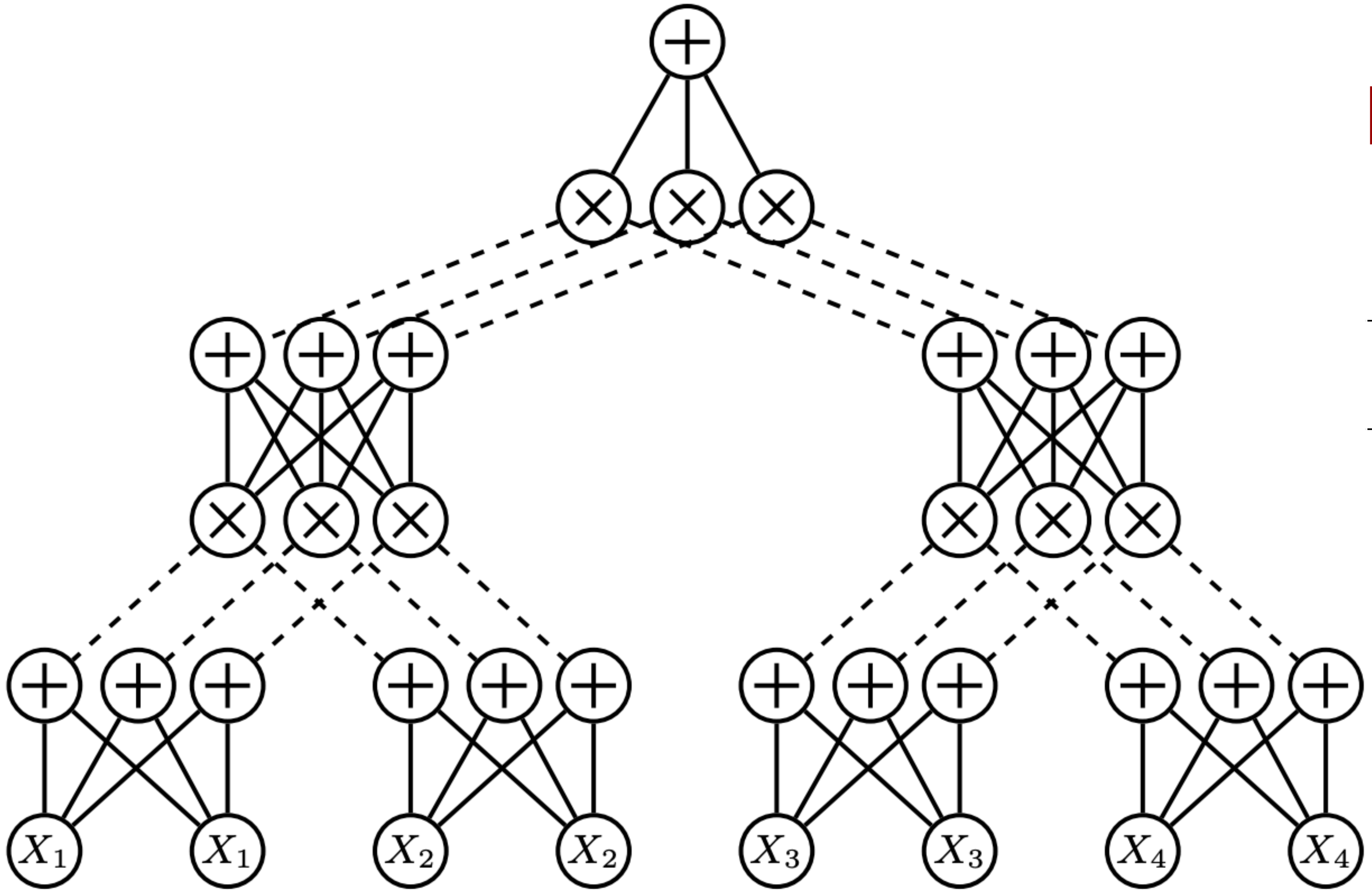


The bigger the network,
the more expressive the model

Large SPNs

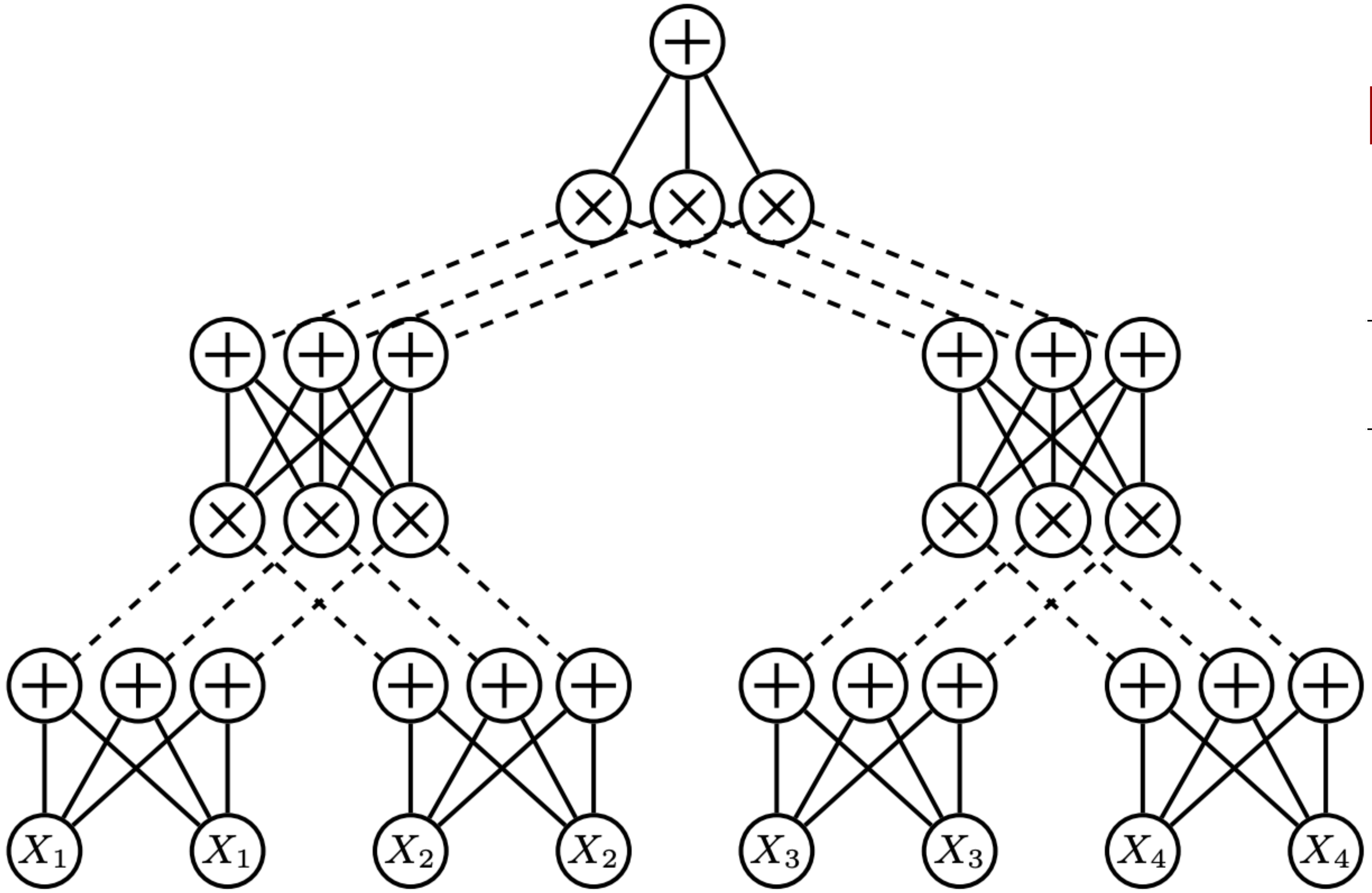


Large SPNs



Regularization Choices

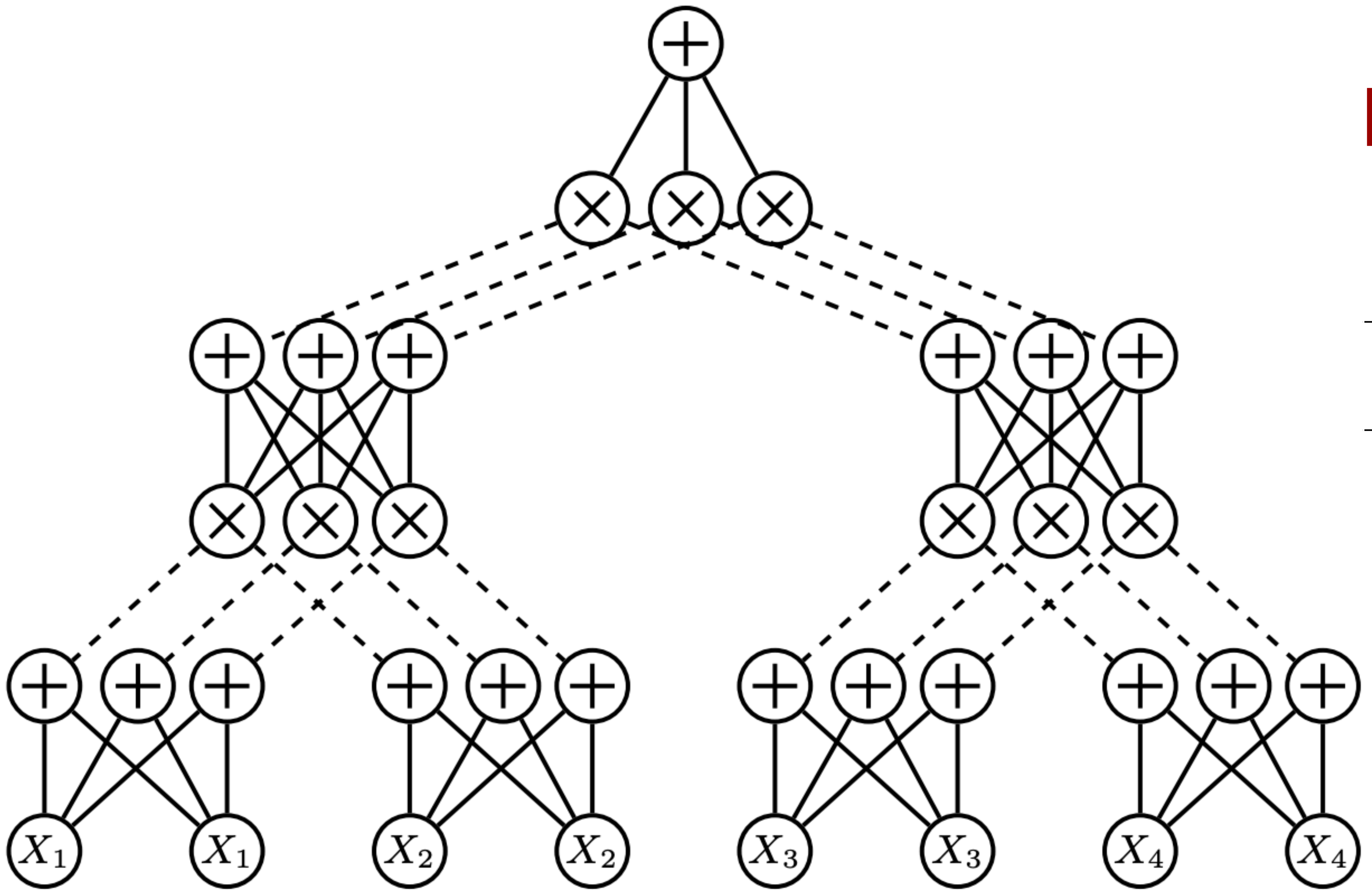
Large SPNs



Regularization Choices

Dropout	Discriminative only

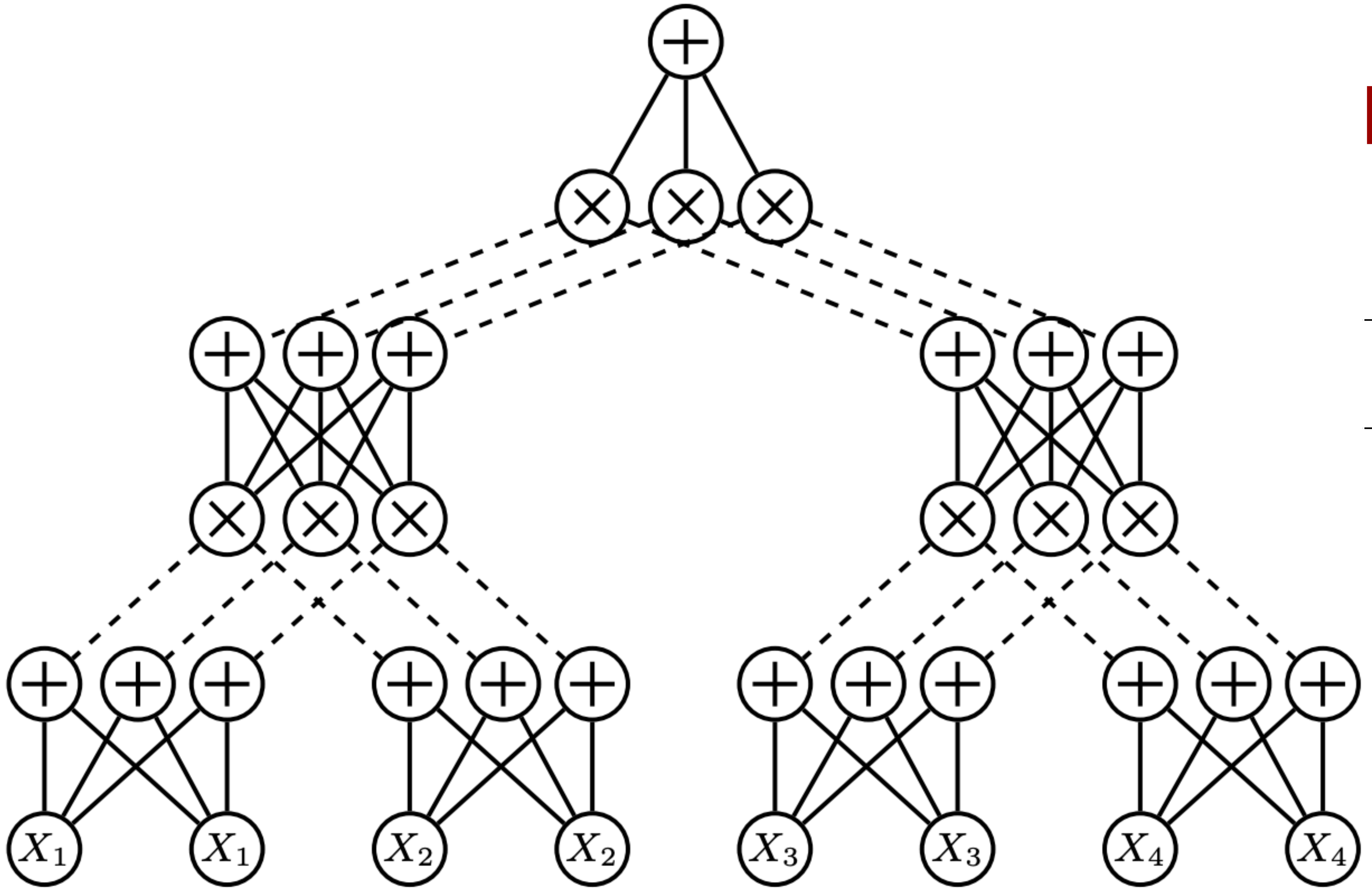
Large SPNs



Regularization Choices

Dropout	Discriminative only
Weight Decay	Many parameters

Large SPNs

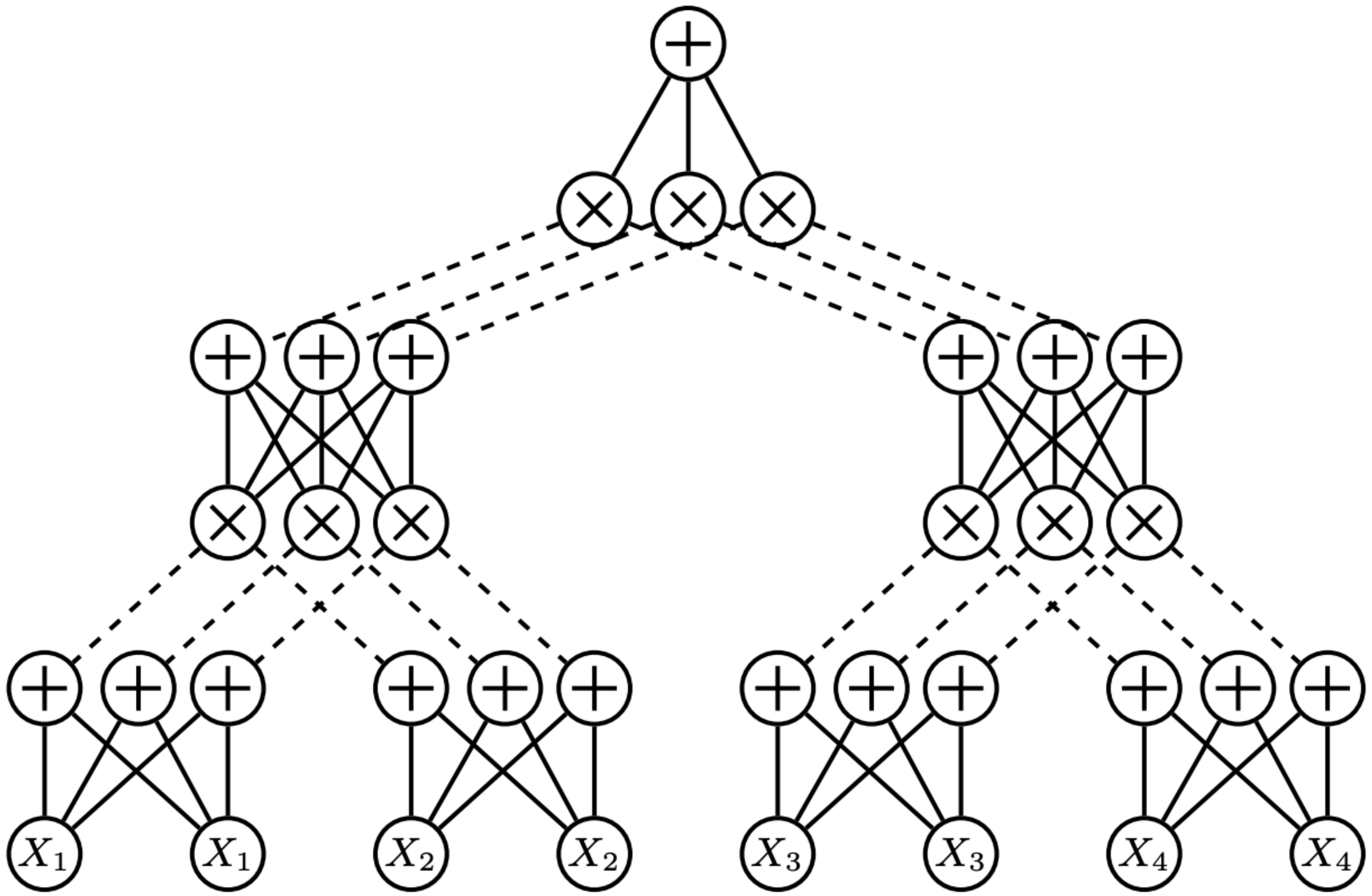


Regularization Choices

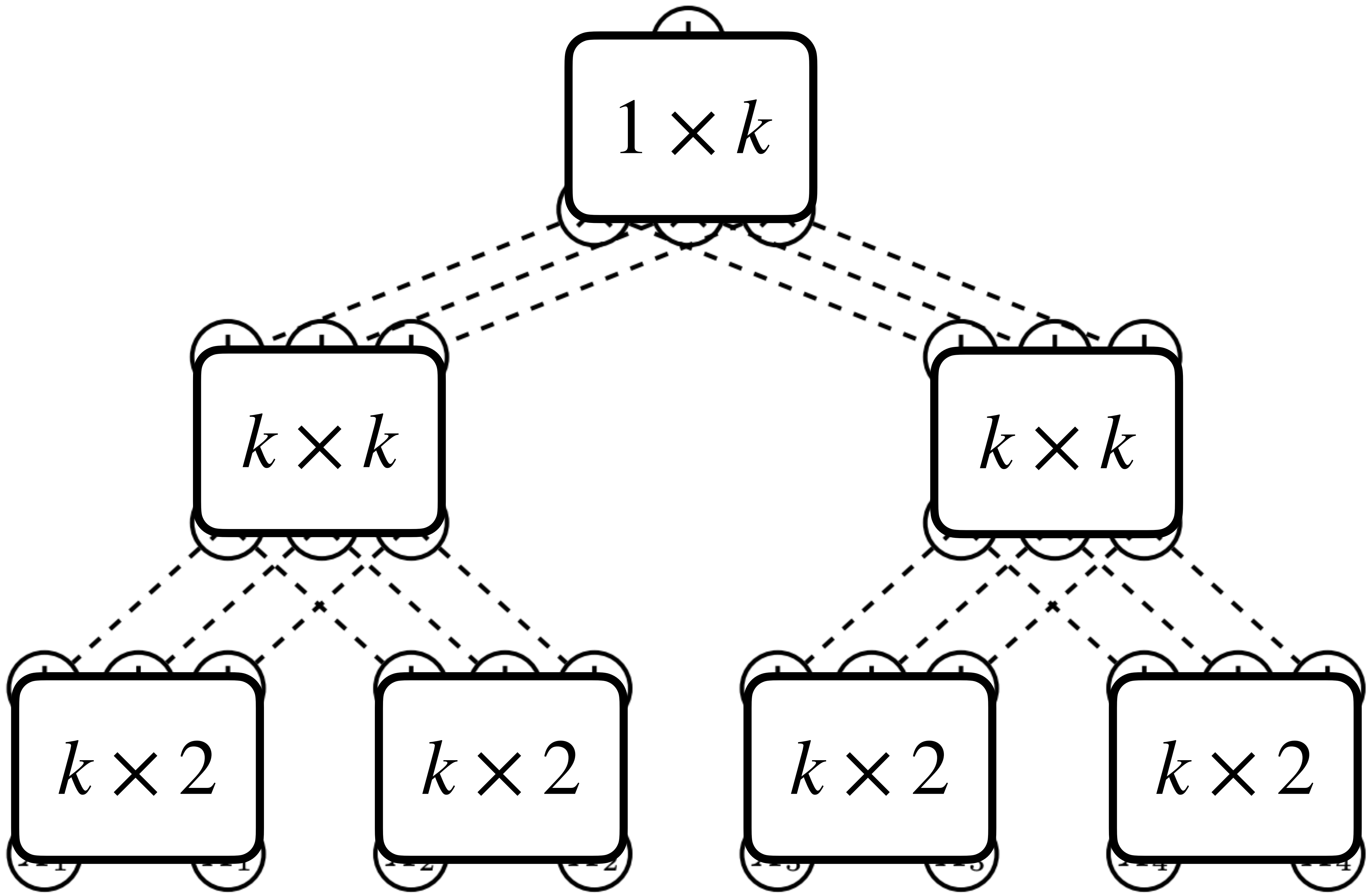
Dropout	Discriminative only
Weight Decay	Many parameters
HyperSPN	Few parameters Memory efficient Better generalization

our proposal

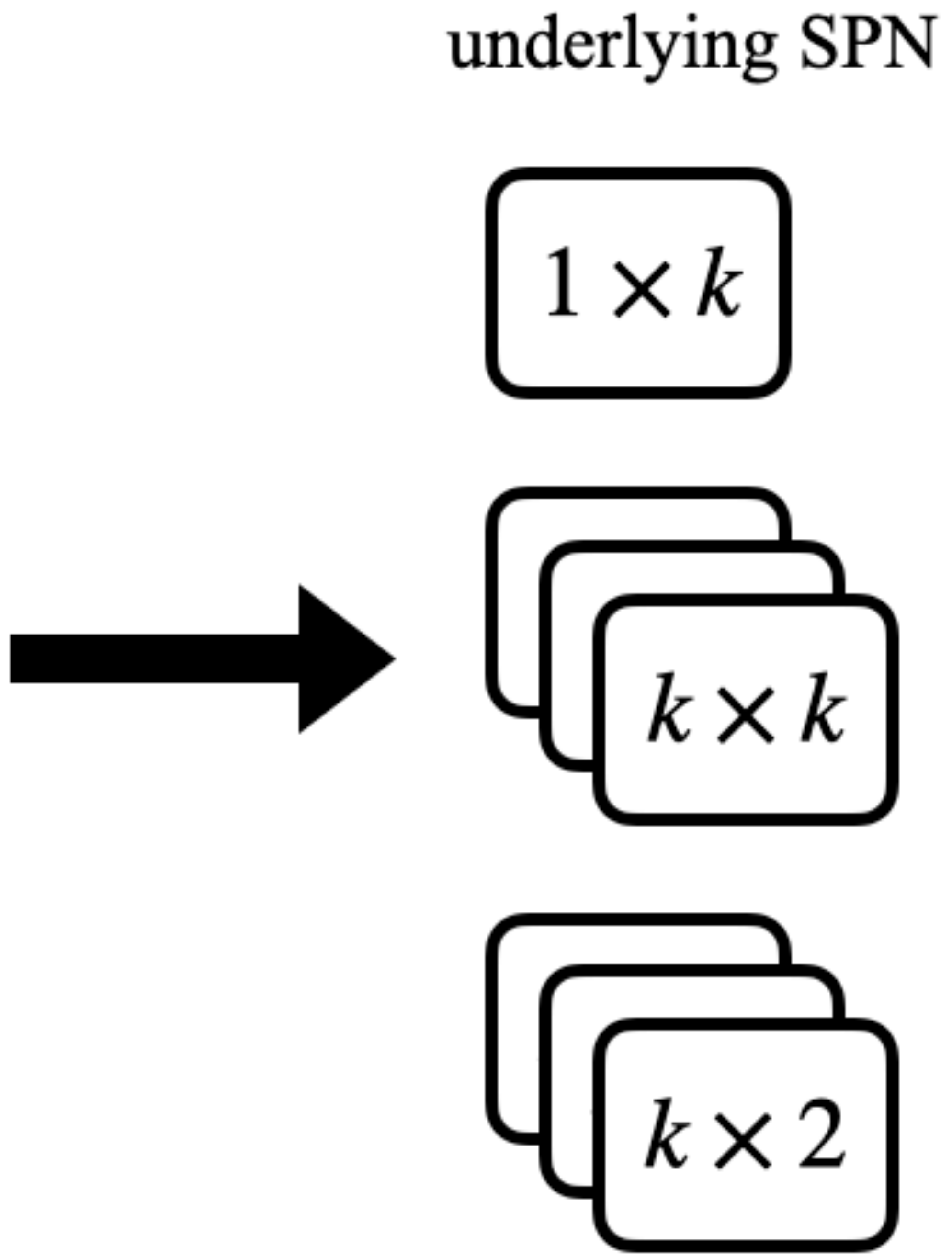
HyperSPN



HyperSPN



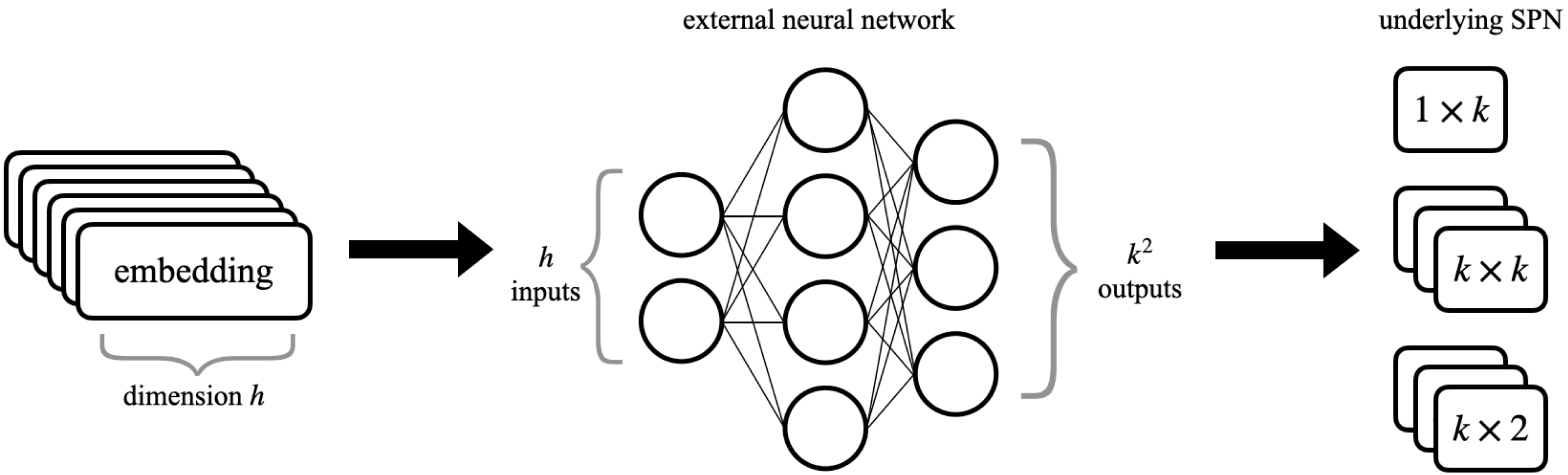
HyperSPN



HyperSPN

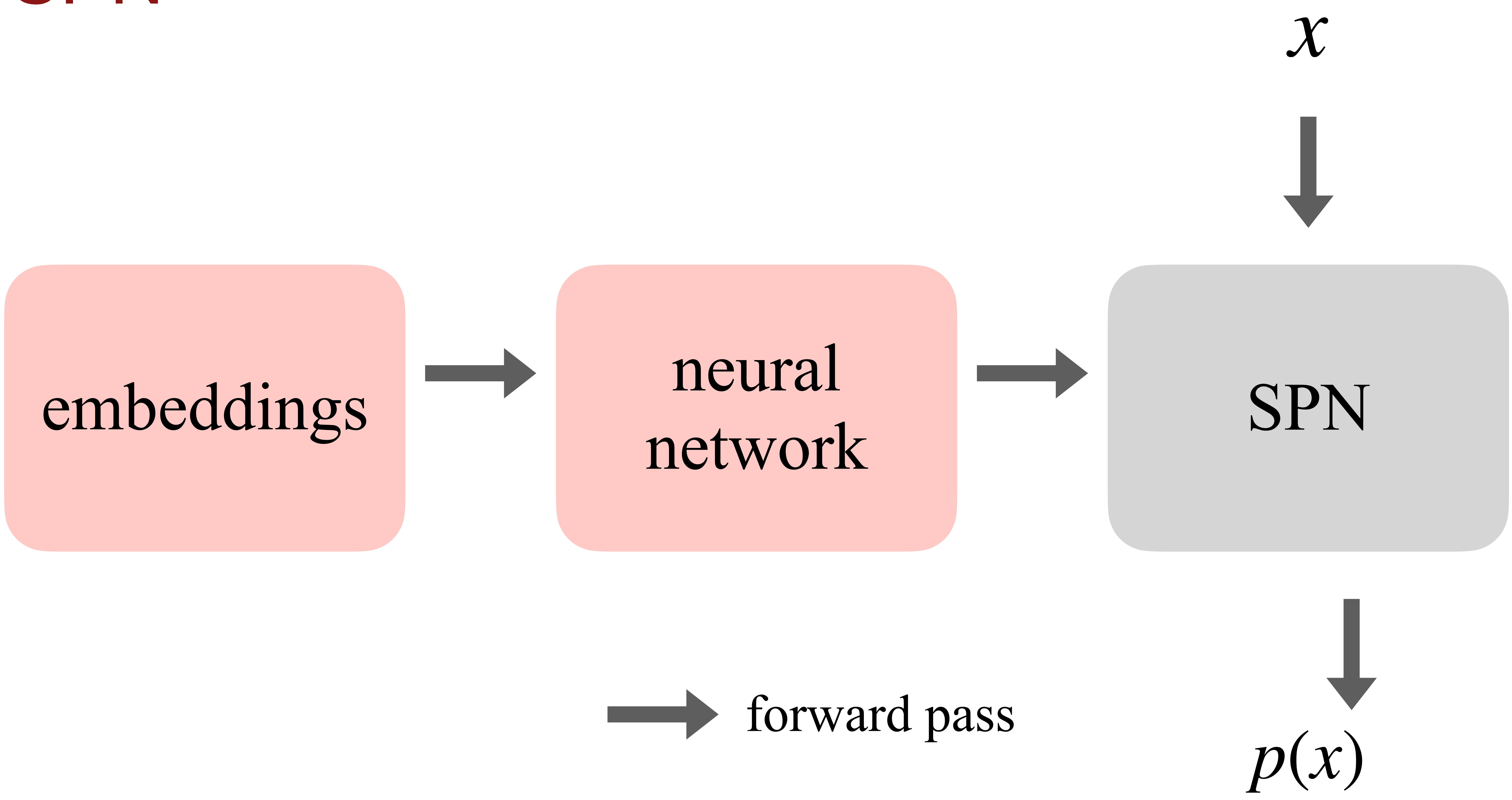

embedding
dimension h
underlying SPN
$1 \times k$
$k \times k$
$k \times 2$

HyperSPN

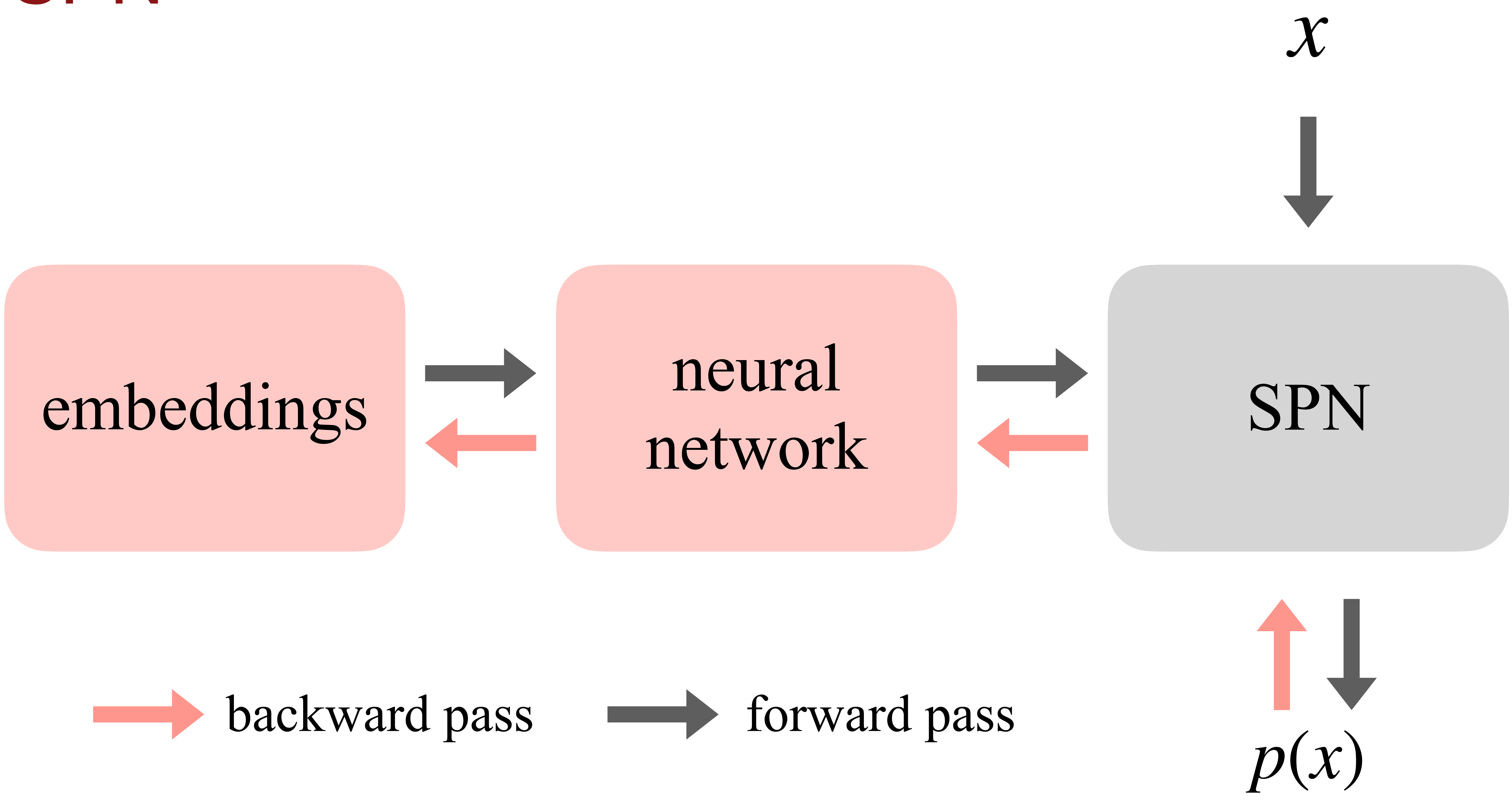


embedding
dimension h
h inputs
external neural network
k^2 outputs
underlying SPN
$1 \times k$
$k \times k$
$k \times 2$

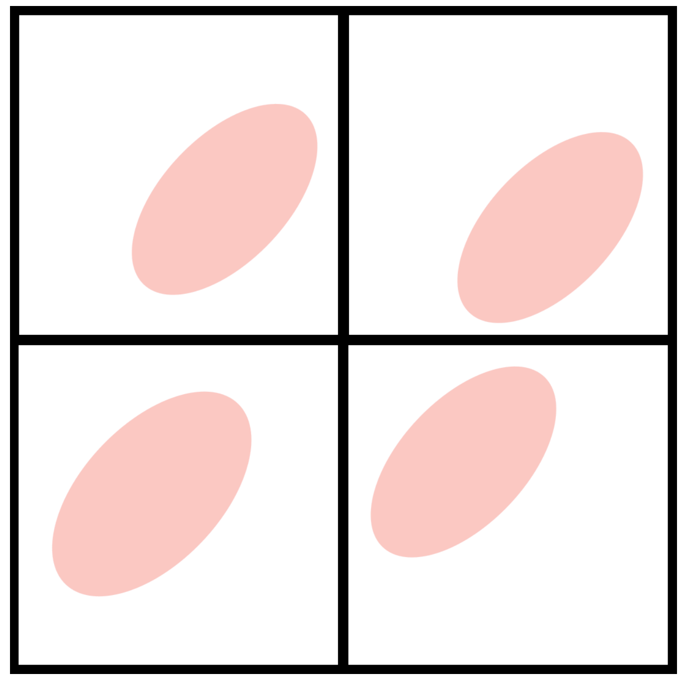
HyperSPN



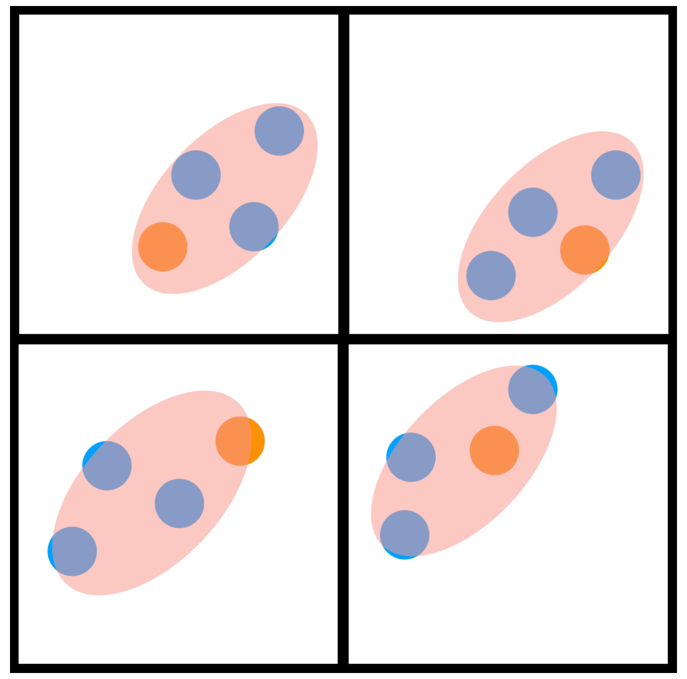
HyperSPN


x
embeddings
neural network
SPN
backward pass
forward pass
$p(x)$

HyperSPN

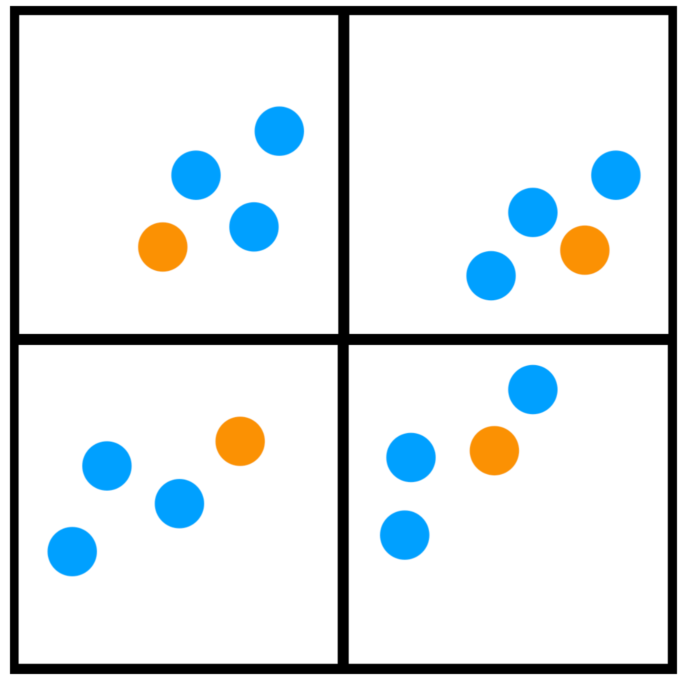


HyperSPN

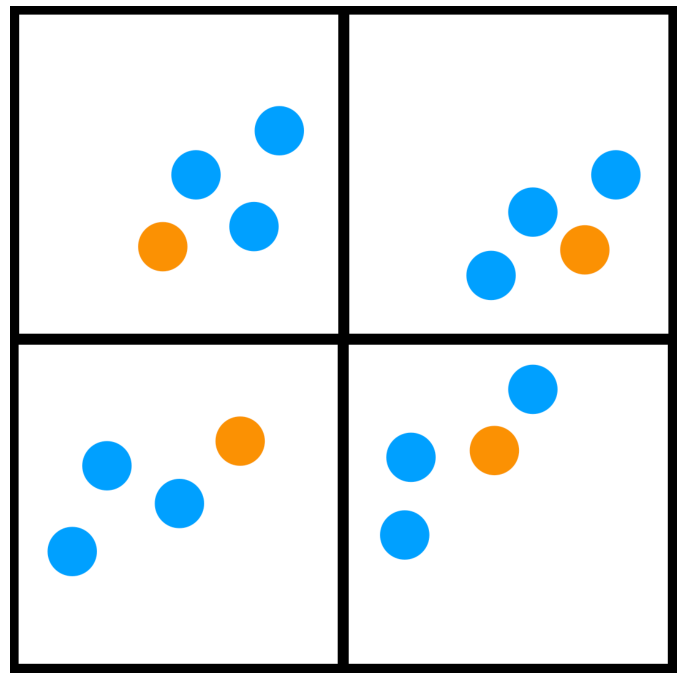


● train ● test

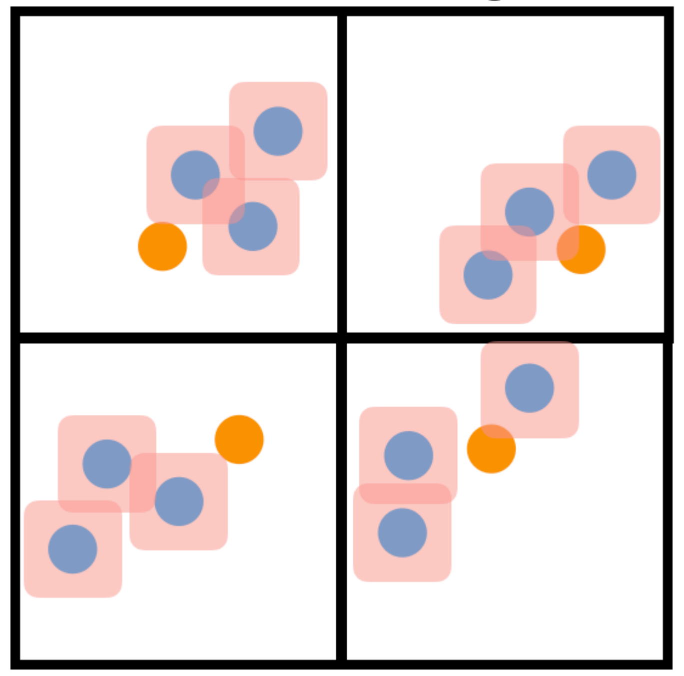
HyperSPN



HyperSPN



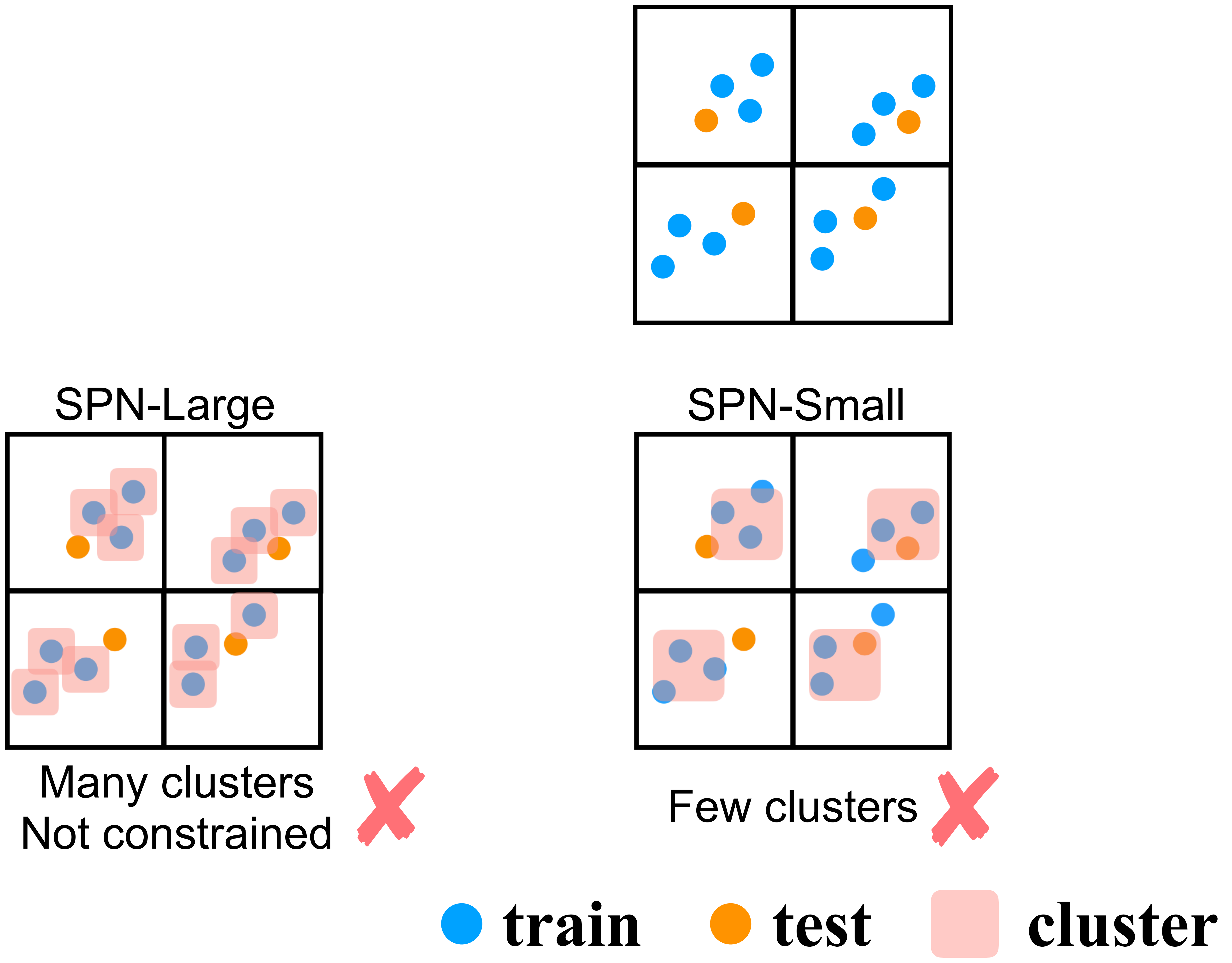
SPN-Large



Many clusters
Not constrained

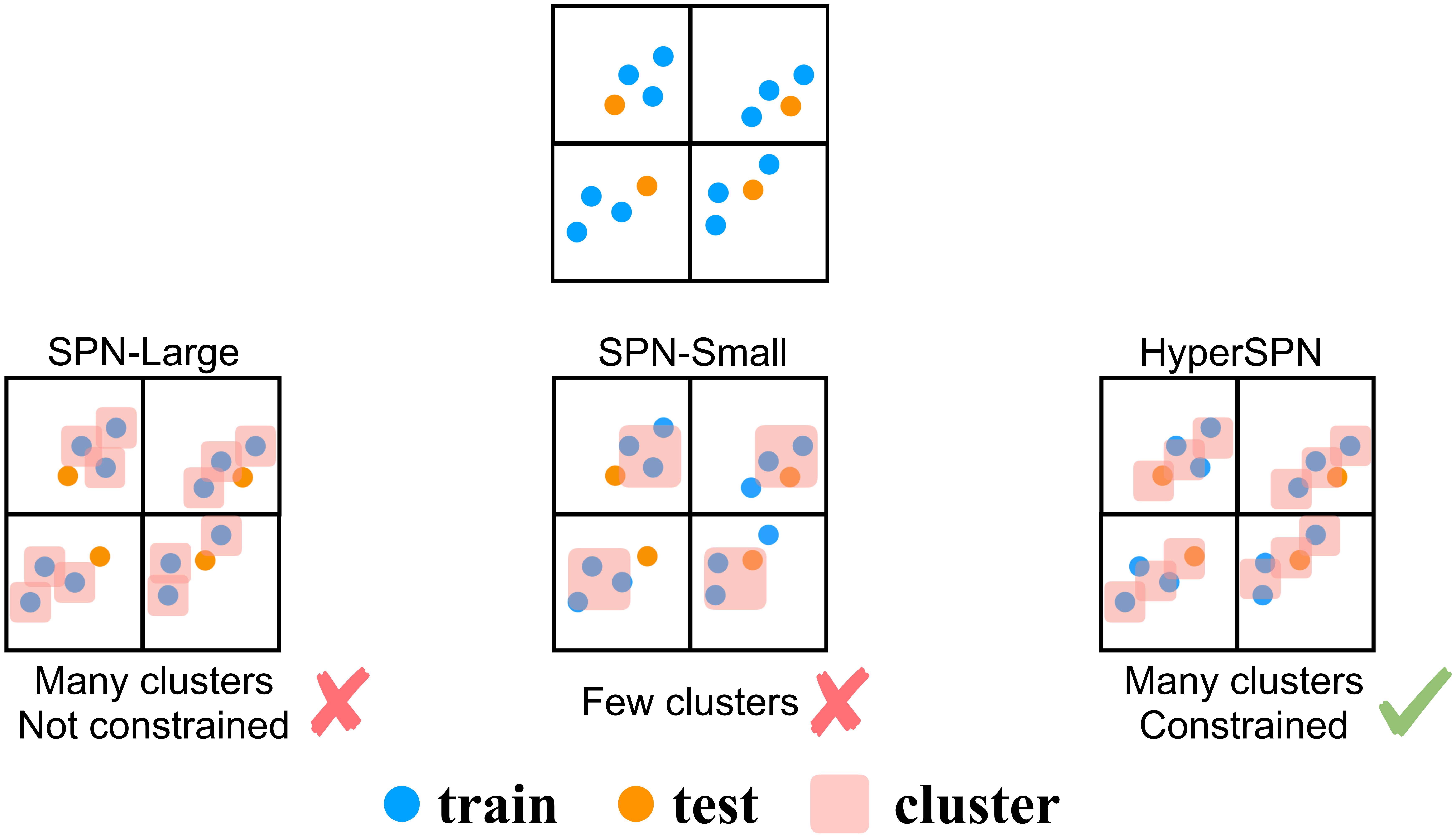
train test cluster

HyperSPN



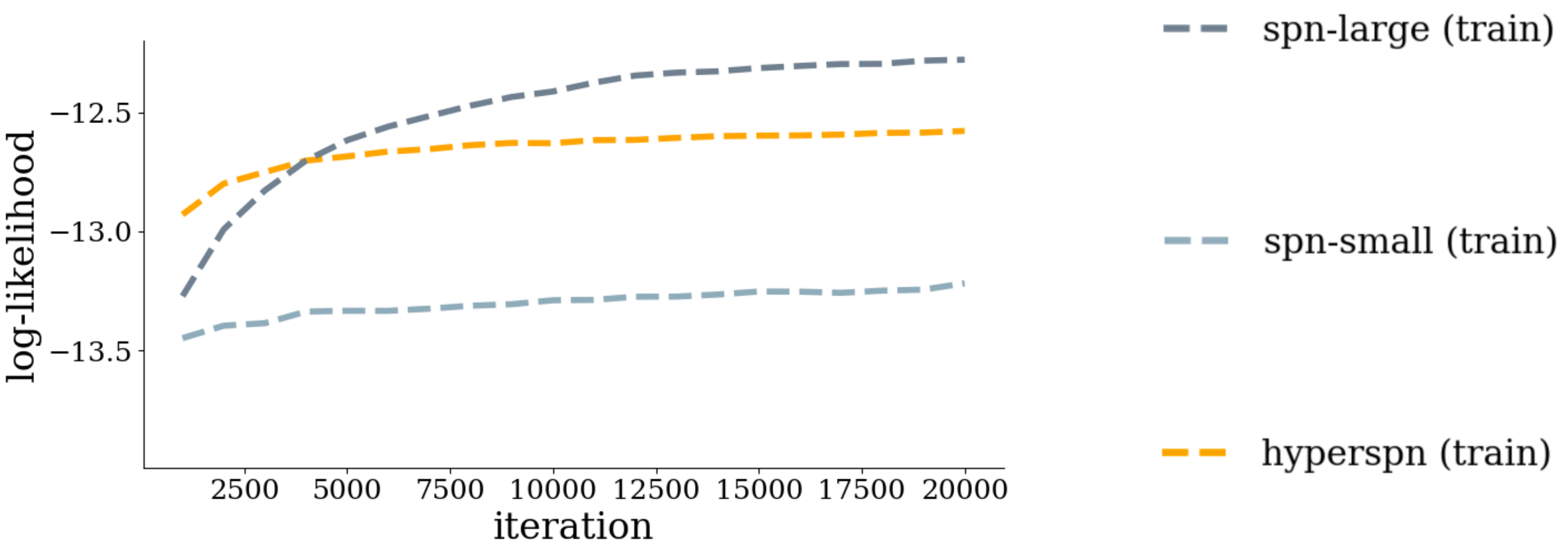
SPN-Large
Many clusters
Not constrained
SPN-Small
Few clusters
train
test
cluster

HyperSPN

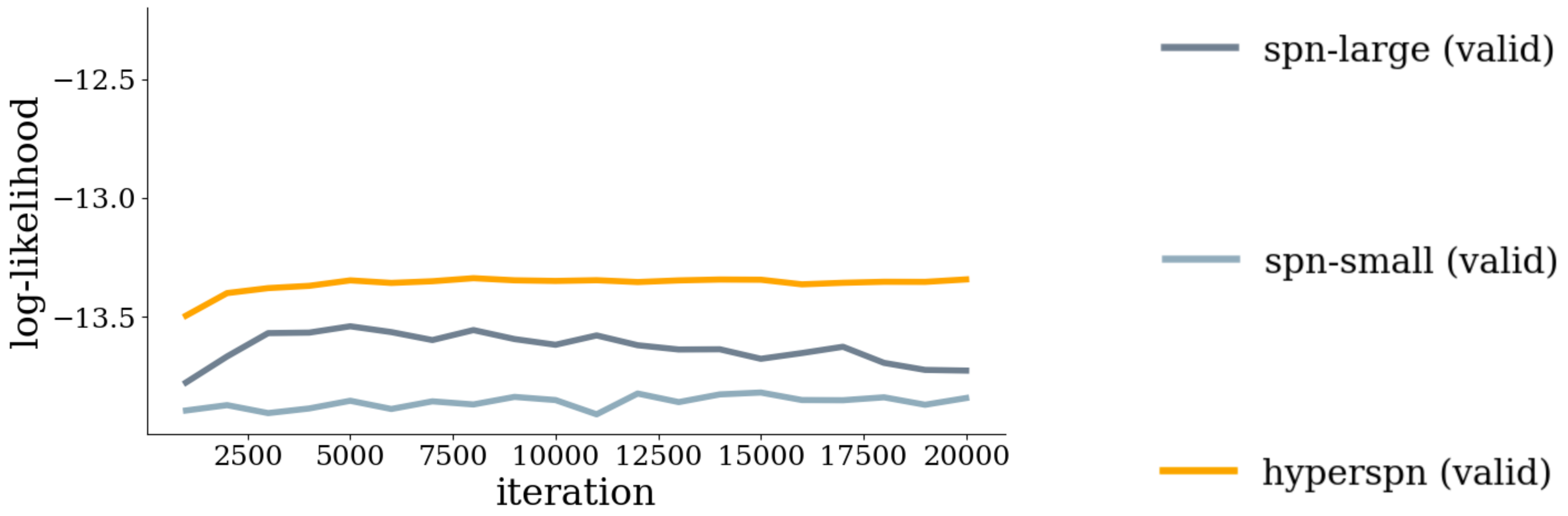


SPN-Large
SPN-Small
HyperSPN
Many clusters
Not constrained
Few clusters
Many clusters
Constrained
train
test
cluster

Better Generalization



Better Generalization



Results

Name	Variables
NLTCS	16
MSNBC	17
KDDCup2k	64
Plants	69
Audio	100
Jester	100
Netflix	100
Accidents	111
Retail	135
Pumsb-star	163
DNA	180
Kosarek	190
MSWeb	294
Book	500
EachMovie	500
WebKB	839
Reuters-52	889
20Newsgrp	910
BBC	1058
Ad	1556

Results

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

Results

better log-LH

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

Results

better log-LH

fewer # params

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

Results

better log-LH

fewer # params

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

HyperSPNs

- regularize by encoding parameters with small NN

Results

better log-LH

fewer # params

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

HyperSPNs

- regularize by encoding parameters with small NN
- better generalization
- more memory efficient
- keeps tractability of SPNs

Results

better log-LH

fewer # params

Name	Variables	Adam Weight Decay Log-LH	Adam Weight Decay # Params	Adam HyperSPN Log-LH	Adam HyperSPN # Params
NLTCS	16	-6.02	40050	**-6.01**	9115
MSNBC	17	**-6.05**	42550	**-6.05**	9615
KDDCup2k	64	-2.14	160050	**-2.13**	33115
Plants	69	-13.36	172550	**-13.26**	35615
Audio	100	-40.18	250050	**-39.83**	51115
Jester	100	-52.98	250050	**-52.75**	51115
Netflix	100	-57.15	250050	**-56.74**	51115
Accidents	111	-36.09	277550	**-35.36**	56615
Retail	135	-10.91	337550	**-10.89**	68615
Pumsb-star	163	-31.76	407550	**-30.79**	82615
DNA	180	**-98.41**	450050	-98.49	91115
Kosarek	190	-10.93	475050	**-10.89**	96115
MSWeb	294	-10.40	735050	**-9.90**	148115
Book	500	-35.01	1250050	**-34.90**	251115
EachMovie	500	-52.99	1250050	**-51.32**	251115
WebKB	839	-159.91	2097550	**-158.60**	420615
Reuters-52	889	-90.14	2222550	**-85.65**	445615
20Newsgrp	910	-154.37	2275050	**-152.49**	456115
BBC	1058	-262.01	2645050	**-254.44**	530115
Ad	1556	-52.23	3890050	**-28.25**	779115

HyperSPNs

- regularize by encoding parameters with small NN
- better generalization
- more memory efficient
- keeps tractability of SPNs

Paper / Code: