Subtractive Mixture Models via Squaring: Representation and Learning

Lorenzo Loconte
University of Edinburgh, UK

Aleksanteri M. Sladek
Aalto University, FI

Stefan Mengel
University of Artois, CNRS, CRIL, FR

Martin Trapp
Aalto University, FI

Arno Solin
Aalto University, FI

Nicolas Gillis
Université de Mons, BE

Antonio Vergari
University of Edinburgh, UK

ICLR 2024 ***Spotlight*** *7-11 May*

Mixture models

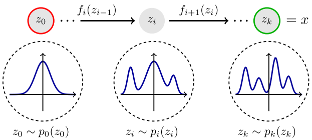
classification
segmentation
clustering
anomaly detection
sequence prediction

"Swiss army knife"
in stats and ML

Bishop and Nasrabadi, "Pattern Recognition and Machine Learning", 2006

Mixture models

classification
segmentation
clustering
anomaly detection
sequence prediction



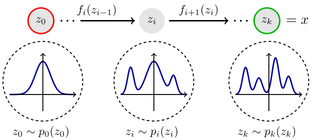
"Swiss army knife" in stats and ML

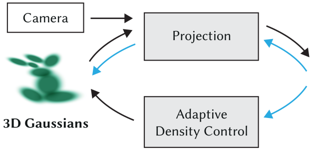
Build more expressive generative models

Bishop and Nasrabadi, "Pattern Recognition and Machine Learning", 2006
Papamakarios et al., "Normalizing flows for probabilistic modeling and inference", 2021
Stimper, Scholkopf, and Hernández-Lobato, "Resampling Base Distributions of Normalizing Flows", 2022

Mixture models

classification
segmentation
clustering
anomaly detection
sequence prediction





"Swiss army knife" in stats and ML

Build more expressive generative models

Fast scene rendering in computer vision

Bishop and Nasrabadi, "Pattern Recognition and Machine Learning", 2006

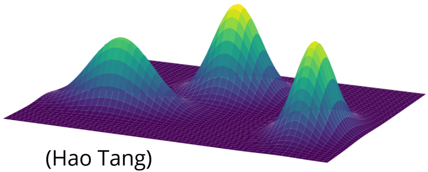
Papamakarios et al., "Normalizing flows for probabilistic modeling and inference", 2021

Stimper, Scholkopf, and Hernández-Lobato, "Resampling Base Distributions of Normalizing Flows", 2022

Kerbl et al., "3D Gaussian Splatting for Real-Time Radiance Field Rendering", 2023

Mixture models

$$p(\mathbf{X}) = \sum_{i=1}^{K} w_i \, p_i(\mathbf{X}) \quad \text{subject to} \quad w_i \geq 0, \quad \sum_{i=1}^{K} w_i = 1$$

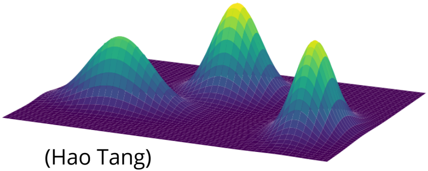


(Hao Tang)

McLachlan, Lee, and Rathnayake, “Finite mixture models”, 2019

Mixture models

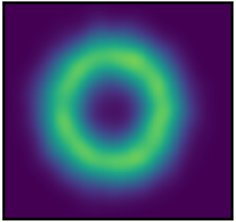
$$p(\mathbf{X}) = \sum_{i=1}^{K} w_i \, p_i(\mathbf{X}) \qquad \text{subject to} \qquad \boldsymbol{w_i \geq 0}, \quad \sum_{i=1}^{K} w_i = 1$$

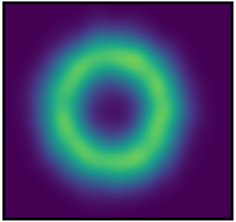


(Hao Tang)

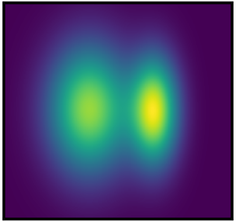
✗ components can only be added together!

McLachlan, Lee, and Rathnayake, “Finite mixture models”, 2019

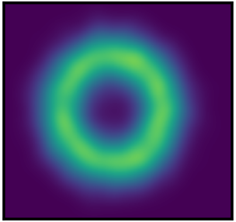

Ground Truth



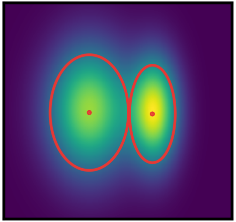
Ground Truth



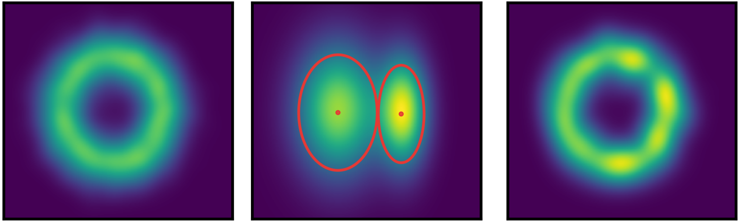
$w_1\mathcal{N}_1 + w_2\mathcal{N}_2$



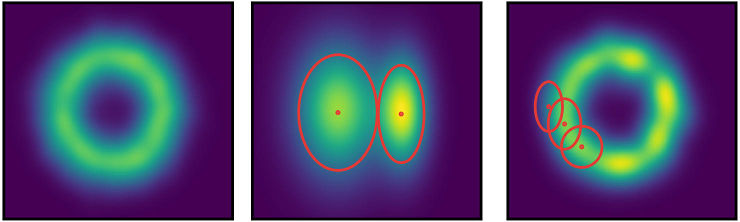
Ground Truth



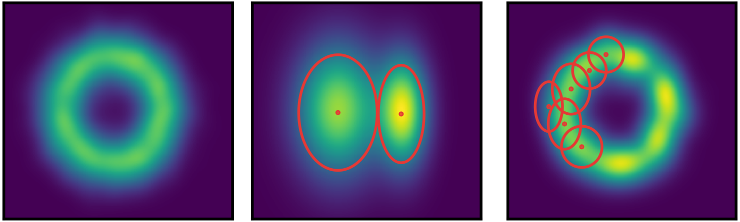
$w_1\mathcal{N}_1 + w_2\mathcal{N}_2$



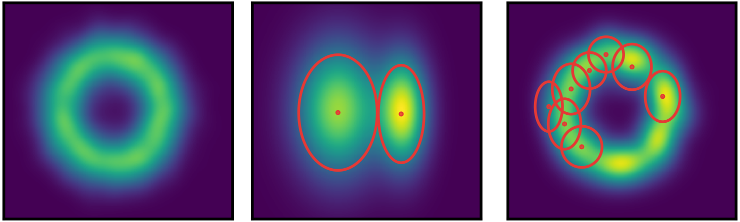
Ground Truth $w_1\mathcal{N}_1 + w_2\mathcal{N}_2$ $w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$



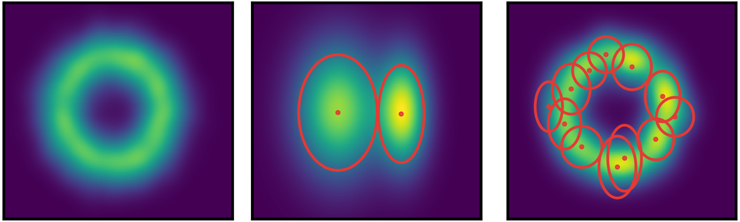
Ground Truth | $w_1\mathcal{N}_1 + w_2\mathcal{N}_2$ | $w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$



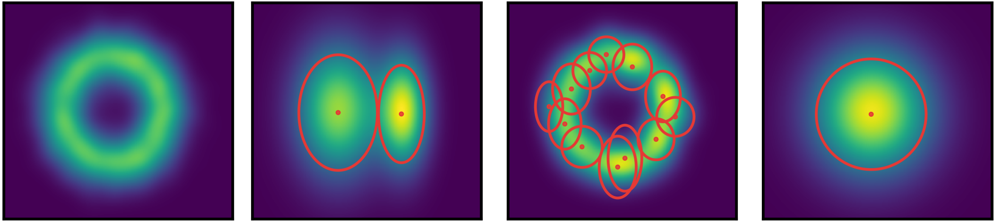
Ground Truth $\quad w_1\mathcal{N}_1 + w_2\mathcal{N}_2 \quad w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$



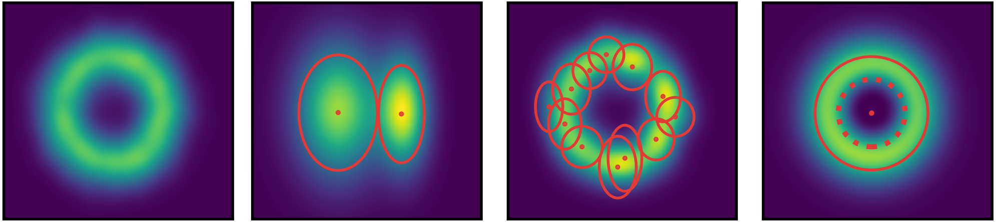
Ground Truth $\quad w_1\mathcal{N}_1 + w_2\mathcal{N}_2 \quad w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$



Ground Truth | $w_1\mathcal{N}_1 + w_2\mathcal{N}_2$ | $w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$



Ground Truth | $w_1\mathcal{N}_1 + w_2\mathcal{N}_2$ | $w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$ | $\mathcal{N}_1$



Ground Truth

$w_1\mathcal{N}_1 + w_2\mathcal{N}_2$

$w_1\mathcal{N}_1 + \cdots w_K\mathcal{N}_K$

$\mathcal{N}_1 - w_2\mathcal{N}_2$

Far fewer components with subtractions

Contributions

I *How to learn subtractive mixture models?*

Contributions

I — *How to learn subtractive mixture models?*

II — *How much more expressive subtractive mixtures are?*

Contributions

I *How to learn subtractive mixture models?*

II *How much more expressive subtractive mixtures are?*

III *What is the relationship with other probabilistic models?*

Squaring mixtures

$$p(\mathbf{X}) \propto \sum_{i=1}^{K} \boldsymbol{w_i}\, p_i(\mathbf{X}), \qquad \boldsymbol{w_i} \in \mathbb{R}$$

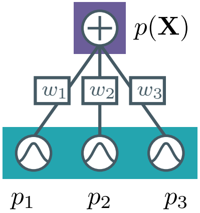
How to ensure $p(\mathbf{X})$ is positive?

Squaring mixtures

$$p(\mathbf{X}) \propto \left(\sum_{i=1}^{K} \boldsymbol{w_i}\, p_i(\mathbf{X})\right)^2 = \sum_{i=1}^{K}\sum_{j=1}^{K} \boldsymbol{w_i}\boldsymbol{w_j}\, p_i(\mathbf{X})p_j(\mathbf{X}), \quad \boldsymbol{w_i} \in \mathbb{R}$$

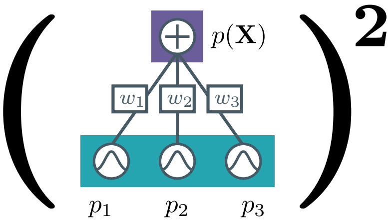
How to ensure $p(\mathbf{X})$ is positive? **By squaring!**

Squaring deep mixtures



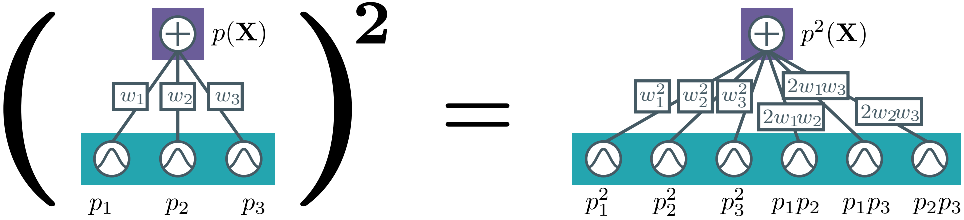
Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020
Vergari et al., "A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference", 2021

Squaring deep mixtures



Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020
Vergari et al., "A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference", 2021

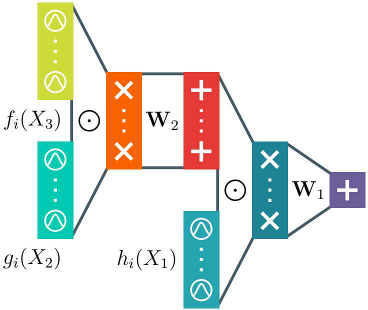
Squaring deep mixtures



Choi, Vergari, and Broeck, "Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling", 2020

Vergari et al., "A Compositional Atlas of Tractable Circuit Operations for Probabilistic Inference", 2021

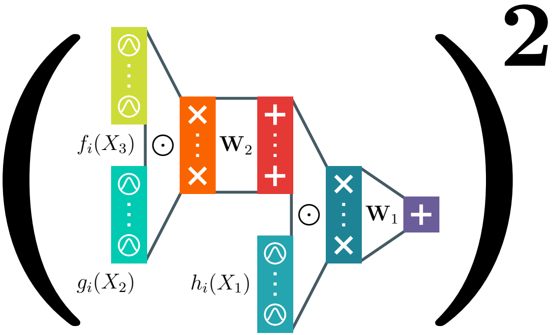
Squaring deep mixtures



Jaini, Poupart, and Yu, “Deep Homogeneous Mixture Models: Representation, Separation, and Approximation”, 2018

Choi, Vergari, and Broeck, “Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling”, 2020

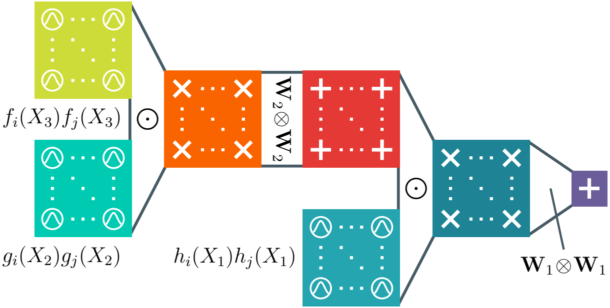
Squaring deep mixtures



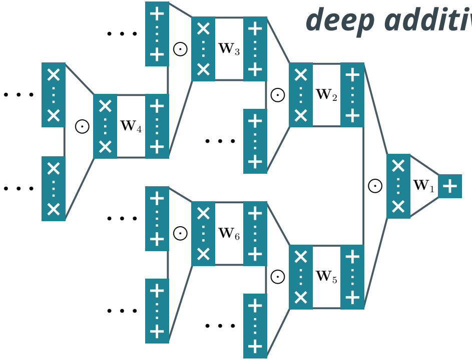
Jaini, Poupart, and Yu, “Deep Homogeneous Mixture Models: Representation, Separation, and Approximation”, 2018

Choi, Vergari, and Broeck, “Probabilistic Circuits: A Unifying Framework for Tractable Probabilistic Modeling”, 2020

Squaring deep mixtures



How much more expressive?

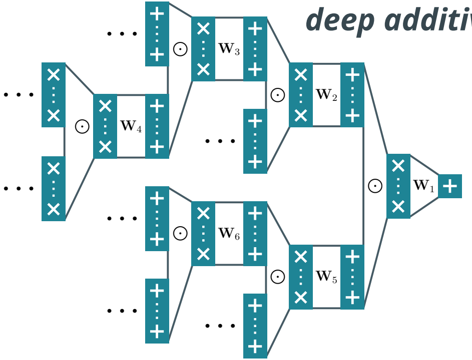


deep additive-only mixtures

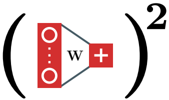
Martens and Medabalimi, “On the expressive efficiency of sum product networks”, 2014

Colnet and Mengel, “A Compilation of Succinctness Results for Arithmetic Circuits”, 2021

How much more expressive?



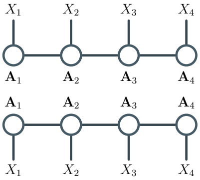
deep additive-only mixtures



squared subtractive mixtures

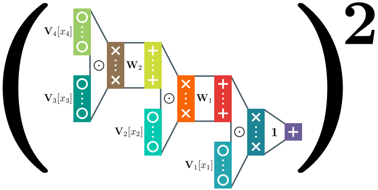
Martens and Medabalimi, “On the expressive efficiency of sum product networks”, 2014
Colnet and Mengel, “A Compilation of Succinctness Results for Arithmetic Circuits”, 2021

Unifying models via squaring



Born machines

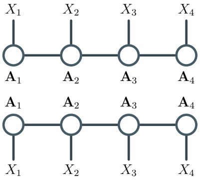
$\Rightarrow$



Glasser et al., "Expressive power of tensor-network factorizations for probabilistic modeling", 2019

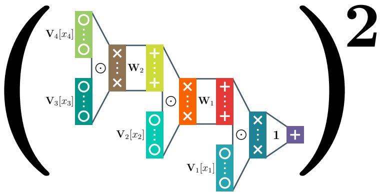
Rudi and Ciliberto, "PSD Representations for Effective Probability Models", 2021

Unifying models via squaring



Born machines

$\Rightarrow$

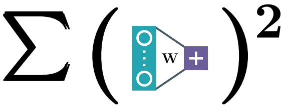


$$p(\mathbf{x}) \propto \kappa(\mathbf{x})^\top \mathbf{A} \kappa(\mathbf{x})$$

with $\mathbf{A} \in \mathbb{R}^{d \times d}$ PSD

Positive Semi-Definite models

$\Rightarrow$



Glasser et al., "Expressive power of tensor-network factorizations for probabilistic modeling", 2019

Rudi and Ciliberto, "PSD Representations for Effective Probability Models", 2021

Takeaways

I *Squared subtractive mixtures ...*

II *... can be much more expressive ...*

III *... and establish a unifying framework*

Poster Session 5

9 May, Halle B, 10:45

Paper

Code

@loreloc_

april

april-tools.github.io