On Finding Modes of Sum-Product Networks

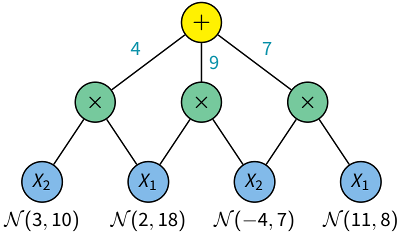
DENIS MAUÁ
joint work with Tiago Madeira and Cassio de Campos

Simons Probabilistic Circuits Workshop – October 20, 2023

University of São Paulo

Sum-Product Network is

- a **tractable univariate distribution** (i.e., we can efficiently find all modes and compute density)
- a **positive combination** of SPNs with same scope
- a **product** of SPNs with disjoint scopes

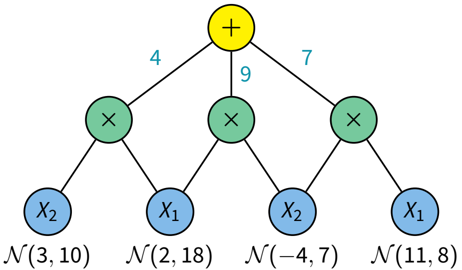


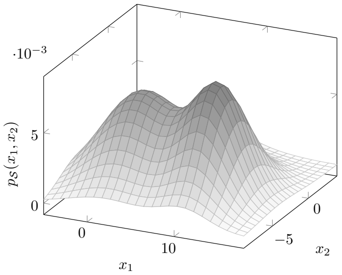
A **Gaussian Sum-Product Network** has only Gaussian distributions at leaves

Modes of a Probability Distribution

Given a **probabilistic circuit** encoding $\Pr(X)$, we say that a configuration x^* is a **mode** if

$$\exists Ne(x^*) \subset \{x^*\} \text{ such that } \Pr(X = x^*) = \max_{x \in Ne(x^*)} \Pr(X = x)$$

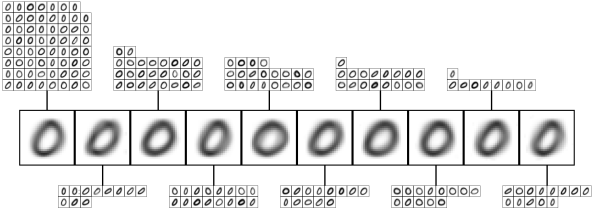




Finding Modes

Motivation

Modes provide good summary of distribution



- Number of modes can be 1, some $1 < k < \infty$ or ∞
- **Related Tasks**:
  - Find all modes
  - Find mode from some point
  - Find most probable modes

Finding A Most Probable Mode (MAP)

- **Given** Probabilistic circuit encoding $\Pr(X)$, find

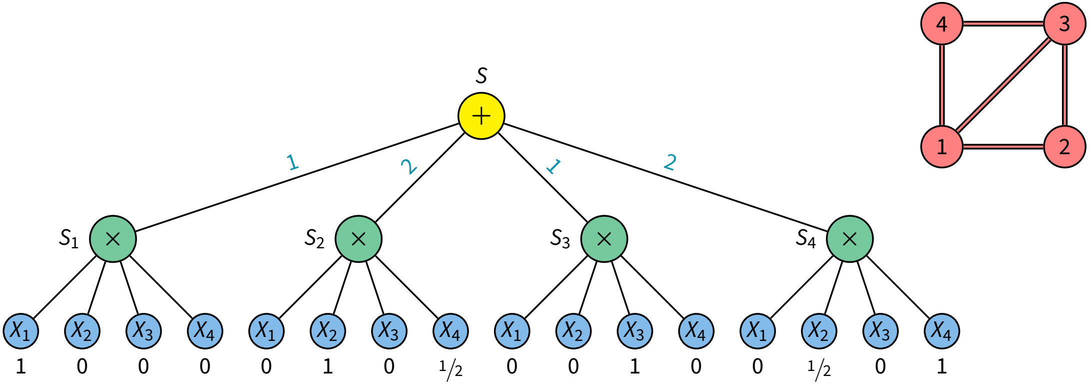
$$\boxed{x^* \in \arg\max \Pr(X = x^*)}$$

- Mode such that $Ne(x^*)$ is entire domain
- **Complexity**
  - NP-hard [de Campos 2011]
  - NP-hard to approximate [Conaty et al. 2017], gets harder with increasing depth
  - Tractable if PC is deterministic [Peharz et al. 2016]
- Applications: Structured prediction, imputation

NP-hardness of Most Probable Mode

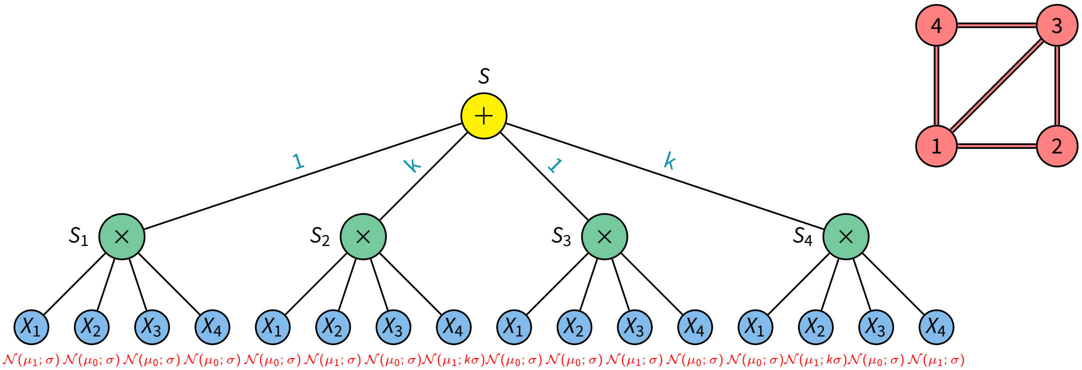
[Conaty et al. 2017]

Proof: Reduction from **maximum independent set**:



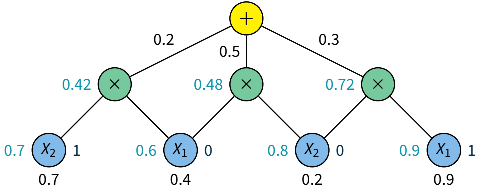
NP-hardness of Most Probable Mode

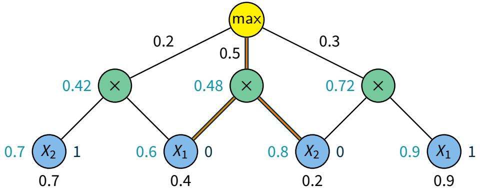
Complexity for **Gaussian Mixture Models** (is this known?)



The Max-Product Algorithm

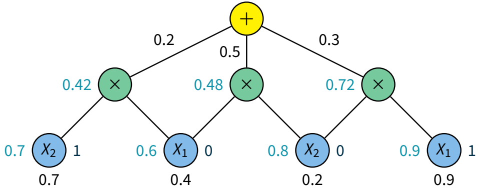
[Poon & Domingos 2011]

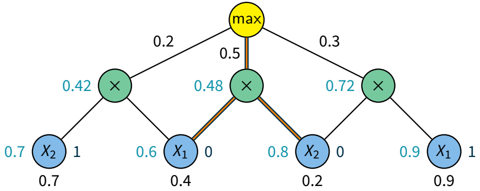


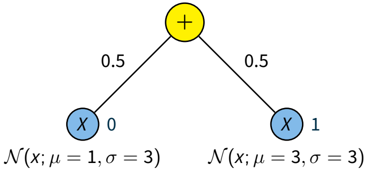


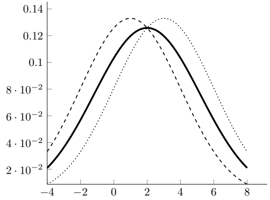
The Max-Product Algorithm

[Poon & Domingos 2011]



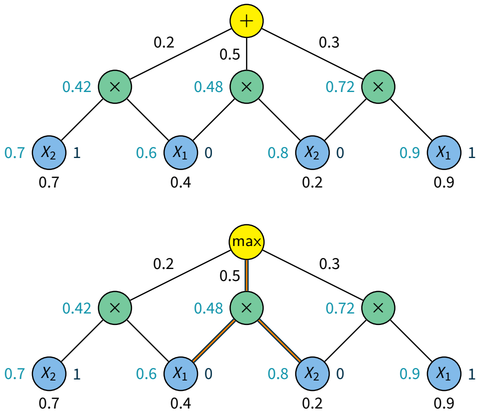






The Max-Product Algorithm

[Conaty et al. 2017]

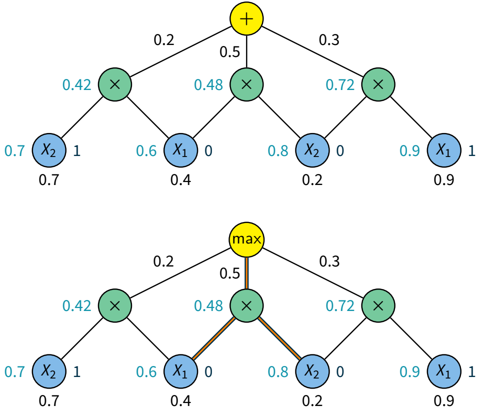


Let $\bar{x}$ be the solution of Max-Product:

$$
\begin{aligned}
\max_x S(x) &\leq \sum_{i=1}^{m} \max_x w_i S_i(x) \\
&\leq m \max_x w_i S_i(x) \\
&\leq m S(\bar{x})
\end{aligned}
$$

The Max-Product Algorithm

[Conaty et al. 2017]



Approximation factor for **discrete SPNs**:

HEIGHT	LOWER BOUND	MAX-PRODUCT
1	1	1
2	n^{ε}	n
≥ 3	$2^{s^{\varepsilon}}$	2^{s}

n: number of internal nodes - 1

s: size of the encoding

If we assume that we can find MAP of mixture of univariate Gaussians, then same upper bounds apply

Continuation Methods

[Pulkkinen et al. 2013]

Significant mode: Mode in a area of high probability (not necessarily a MAP)

- Convolution of Gaussian and GMM has **closed form solution** in the form of unscaled GMM
  - Under certain conditions, convolved dist. is strictly concave
- **Mapping between modes** of convolved distribution and GMM (differential equation)
- Works for GMMs with **negative weights**

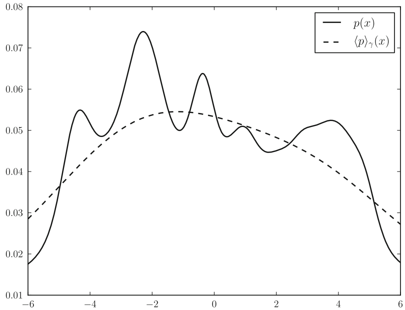


Fig. 1 A Gaussian mixture p and its concave Gaussian convolution $\langle p \rangle_\gamma$ in the interval $[-6, 6]$ with $\gamma = 3$

Finding a Mode From a Starting Point

- **Given** Initial point x_0, find improving solution x_1 with

$$\boxed{\Pr(X = x_1) > \Pr(X = x_0)}$$

  or decide that x_0 is mode

- **Complexity** (on discrete PCs)
  - NP-hard [Bodlaender et al. 2002, Villanueva & Mauá 2020]
  - Tractable if PC is deterministic [Peharz et al. 2016]
  - Tractable if neighborhood is small and fixed (e.g., Hamming distance $<$ constant)

- Applications: Multiple imputation, modal clustering

Finding a Mode From Starting Point

- **Gradient Ascent**: $x_{t+1} = x_t + \eta_t \nabla \Pr(X = x_t)$
- **Gaussian Mean-Shift** [Carreira-Perpiñán 2000]:

$$\boxed{x_{t+1} = \sum_{x \in D} \frac{x}{\sigma^2} \frac{w_x \mathcal{N}(x_t; \mu = x, \sigma)}{\sum_{x \in D} w_x \mathcal{N}(x_t; \mu = x, \sigma)}}$$

- **Modal EM for GMMs** [Li et al. 2007, Carreira-Perpiñán 2007]:

Expectation:

$$q_k = \frac{w_k \mathcal{N}(x_t; \mu_k, \sigma_k)}{\sum_k w_k \mathcal{N}(x_t; \mu_k, \sigma_k)}$$

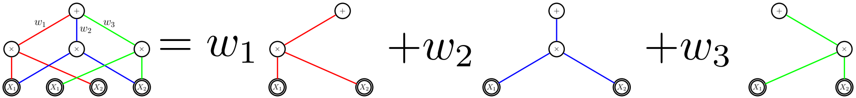
Maximization:

$$x_{t+1} = \arg\max_x \sum_k q_k \log \mathcal{N}(x; \mu_k, \sigma_k)$$

Induced Circuits

[Zhao et al. 2015]

An **induced circuit** T of a circuit C is a subcircuit obtained by selecting exactly one child (input) for each sum node.



- Each induced circuit T_k is a tree/product distribution and

$$C(x) = \sum_k T_k(x)$$

Modal EM for Gaussian SPNs

[Madeira & Mauá, 2022]

Let $T_k(x) = w_k \prod_\ell \mathcal{N}(x[\ell]; \mu_{k,\ell}, \sigma_{k,\ell})$ be the distribution of the k-th induced circuit of C

The Modal EM on circuit C satisfies

Expectation:

$$q_k = \frac{w_k \prod_\ell \mathcal{N}(x_t; \mu_{k,\ell}, \sigma_{k,\ell})}{\sum_k w_k \prod_\ell \mathcal{N}(x_t; \mu_{k,\ell}, \sigma_{k,\ell})}$$

Maximization:

$$x_{t+1}[j] = \arg\max_x \sum_k q_k \sum_\ell \log \mathcal{N}(x[\ell]; \mu_\ell, \sigma_\ell)$$

$$= \sum_k \frac{\mu_{k,j}}{\sigma_{k,j}^2} \frac{T_k(x)}{\sum_k \sigma_{k,\ell}^{-2} T_k(x)}$$

Modal EM for GSPNs I

Input: Gaussian Sum-Product Network C with m nodes and configuration $x \in \mathbb{R}^n$

1. Create sparse arrays N and D of size m-by-n

2. Visit each node u from the leaves (inputs) to the root (output) and compute:

Modal EM for GSPNs II

- **Leaf Node** with scope X_i and parameters μ and σ

$$N_{u,i} = \frac{\mu}{\sigma^2}\mathcal{N}(x[i];\mu,\sigma) \qquad D_{u,i} = \frac{1}{\sigma^2}\mathcal{N}(x[i];\mu,\sigma)$$

$$N_{u,j} = \mathcal{N}(x[i];\mu,\sigma) \qquad D_{u,j} = \mathcal{N}(x[i];\mu,\sigma) \qquad [\forall X_j \neq X_i]$$

Modal EM for GSPNs III

- **Product Node**

$$N_{u,i} = \prod_{(u,v)} N_{v,i} \qquad D_{u,i} = \prod_{(u,v)} D_{v,i}$$

- **Sum Node**

$$N_{u,i} = \sum_{(u,v)} w_{u,v} N_{v,i} \qquad D_{u,i} = \sum_{(u,v)} w_{u,v} D_{v,i}$$

3. **Return** $x' = N_{r,\cdot} / D_{r,\cdot}$ for root node r

Modal EM for Gaussian SPNs

[Madeira & Mauá, 2022]

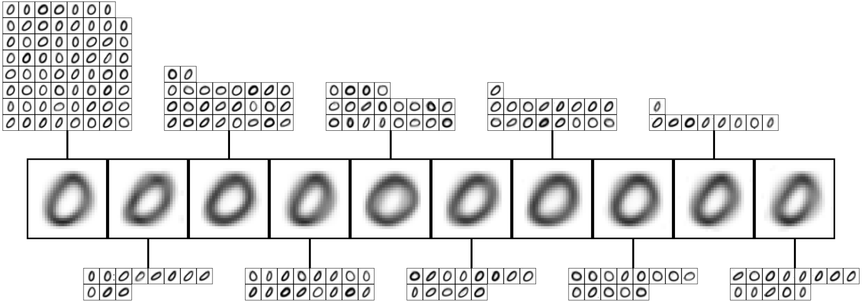
Thm: Modal EM outputs x' such that $C(x') \geq C(x)$ in time $O(mn)$.

Caveat: If $C(x') = C(x)$ then x' might be saddle point, not mode (but seldom occurs in practice).

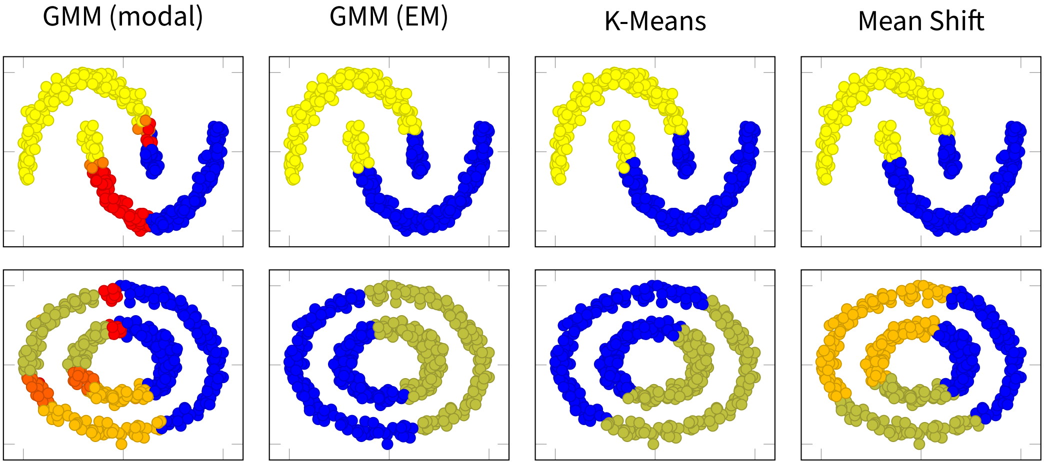
Modal Clustering

[Madeira & Mauá, 2022]

Learn GSPN using LearnSPN, then associate find mode from each datapoint



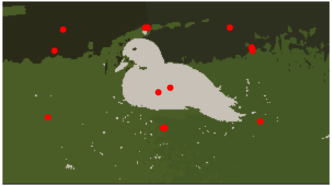
Modal Clustering



Modal Clustering For Image Segmentation[1]

Original

GMM ($k = 6$)

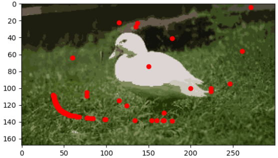


6-means

[1] Experiments/images made by Jonas Gonçalves

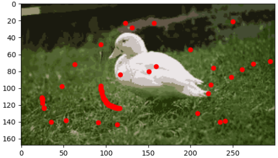
Modal Clustering For Image Segmentation[2]

$k = 20$



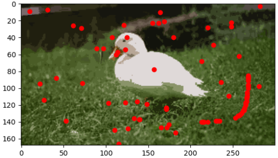
77 modes

$k = 30$



77 modes

$k = 50$



87 modes

[2] Experiments/images made by Jonas Gonçalves

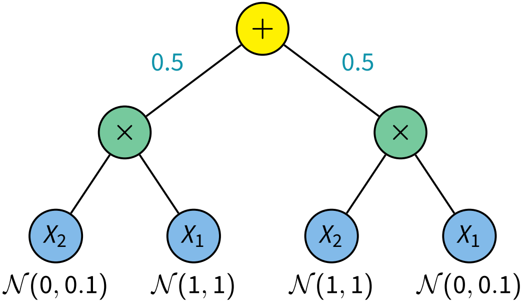
Finding All Modes

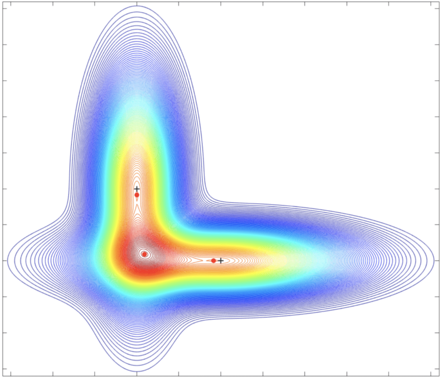
General Search Strategy:

1. Determine/estimate number of modes N
2. While number of modes found $< N$ do:
   2.1 Find point x in candidate region (of mode yet to be found)
   2.2 Run Modal EM from x, check if new mode is found

Finding All Modes

[Améndola et al. 2019]

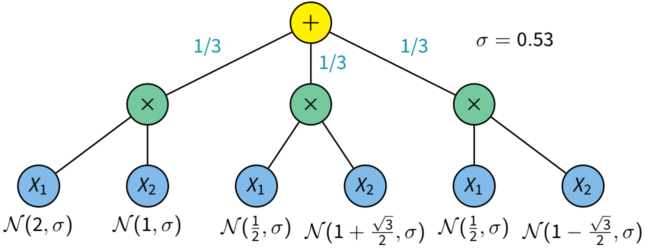


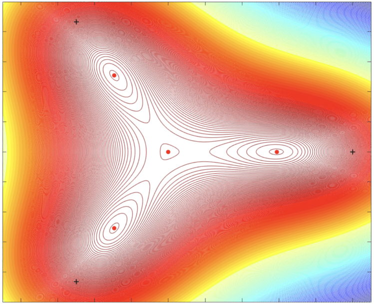


Modes at $x_1 \approx [0, 0.9], x_2 \approx [0.9, 0], x_3 \approx [0.09, 0.09]$ (MAP)

Finding All Modes

[Améndola et al. 2019]



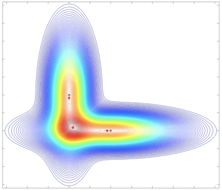


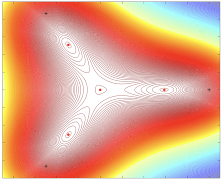
Modes at $x_1 \approx [1.98, 1]$, $x_2 \approx [0.5, 1.8]$, $x_3 \approx [0.5, 0159]$, $x_4 \approx [1, 1]$ (all are MAP)

Finding All Modes

Where to search for modes?

- The modes of **isotropic** (equal variance) GMMs lie in the convex hull of the component means [Améndola et al. 2019]:
  - Use max-product to find starting point
  - Use constrained version to find second point, and so on... (beam search)
- For **non-isotropic**, modes can lie outside convex hull (including MAP)
- **Data-centric**: Use some training data as starting points





Finding All Modes

When to stop searching for modes:

- Unidimensional mixtures have at most number of components [Carreira-Perpiñán & Williams 2003]
- There are n-dimensional GMM with 2 components and $n + 1$ modes [Ray & Ren 2012]
- For any $k, n \geq 2$, there are n-dimensional GMMs with k components and $O(k^n)$ modes [Améndola et al. 2019]
  - Number of stationary points is at most $O(2^{n+k^2} n^k)$
  - For $k = 100$ and $n = 50$ that gives us roughly 10^{29} possible modes
- **Practice**: Modes appear near components (which can exponentially many in GSPNs)

Summary

- Modes give a good summary of model, with useful applications
- Finding Modes of SPNs is a challenging problem
- There is much less work done for the continuous case (and even less in the mixed discrete-continuous case)
- Theoretical open questions about number and location of modes of GSPNs