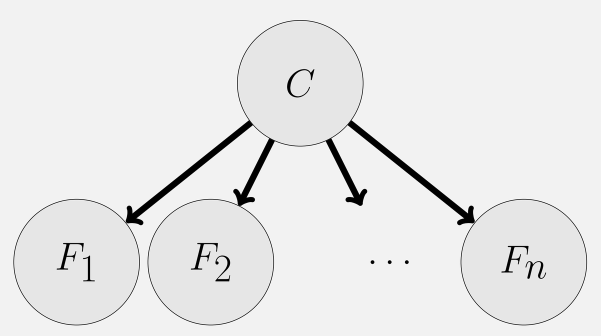
The Multilabel Naive Credal Classifier

Alessandro Antonucci and Giorgio Corani
Dalle Molle Institute for Artificial Intelligence (IDSIA) - Lugano (Switzerland)

Istituto Dalle Molle di studi sull'intelligenza artificiale
USI/SUPSI IDSIA

Naive Credal Classifier (single-label)



Class $C = c \in \mathcal{C} = (c_1, \dots, c_n)$ Feats $\boldsymbol{F} := (F_1, \dots, F_m)$

Naive topology $\Rightarrow P(c, \boldsymbol{f}) = P(c) \prod P(f_i|c)$

Precise learning with single Dirichlet prior

$P(\theta_c) \propto \theta_c^{st(c)-1}$ with $\sum_c t(c) = 1$ and $P(c) = \theta_c$

$P(c) = \frac{n(c)+st(c)}{N+s}$ $\quad P(f_i|c) = \frac{n(c,f_i)+st(c,f_i)}{n(c)+st(c)}$

IDM takes all the priors such that: $\sum_c t(c) = 1 \; \sum_{f_i \in \mathcal{F}_i} t(c, f_i) = t(c)$

Given instance $\boldsymbol{f}$, dominance criterion $c' \succ c''$ IFF $P_{\boldsymbol{t}}(c', \boldsymbol{f}) > P_{\boldsymbol{t}}(c'', \boldsymbol{f}) \; \forall \boldsymbol{t}$

Dominance test performed $\forall c', c'' \in \mathcal{C}$, undominated classes in output

Dominance test $\inf_{\boldsymbol{t}} \frac{P(c',\boldsymbol{f})}{P(c'',\boldsymbol{f})} = \inf_{\boldsymbol{t}} \left[\frac{n(c'')+st(c'')}{n(c')+st(c')}\right]^{m-1} \prod_{i=1}^m \frac{n(c',f_i)+st(c',f_i)}{n(c'',f_i)+st(c'',f_i)} > 1$

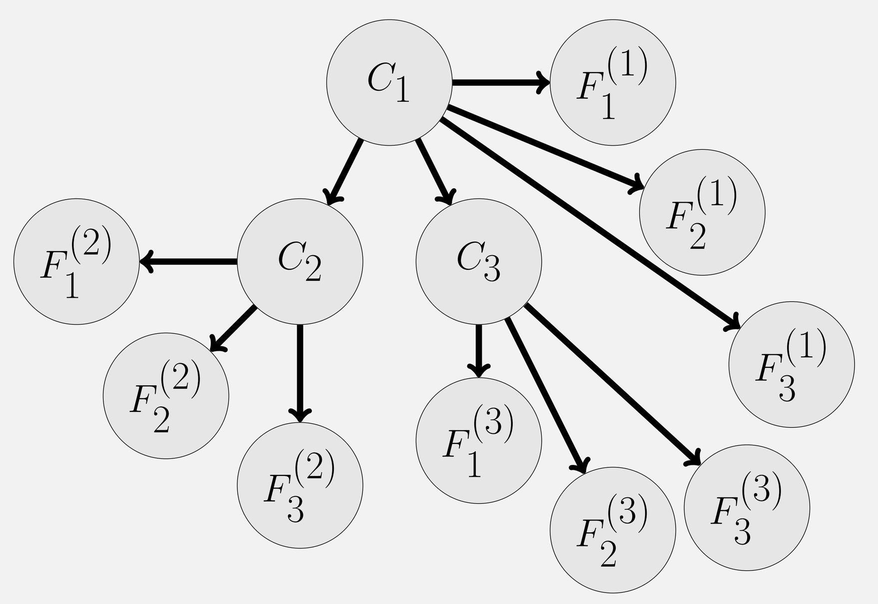
The optimization of the second term can be achieved independently

The test rewrites as $\left[\frac{n(c'')+st(c'')}{n(c')+st(c')}\right]^{m-1} \prod_{i=1}^m \frac{n(c',f_i)}{n(c'',f_i)+st(c'')}$ with $t(c') + t(c'') = 1$

The objective function is a function of a single variable, its logarithmic derivative is a linear fractional variable, and the second derivative is always positive.

The minimization can be efficiently achieved by bracketing (Zaffalon, 2001)

Naive-like Credal Classifier (multi-label)



class labels not exclusive

n Boolean vars $\boldsymbol{C} := (C_1, \dots, C_n)$

Naive-like assumption: given "superclass" C_j other classes are independent

Feature F_j is are common child of all class labels: disconnected by replication (see right)

Factorization: $P(\boldsymbol{c}, \boldsymbol{f}) = P(c_1) \left[\prod_{i=2}^q P(c_i|c_1)\right] \prod_{j=1}^q \prod_{k=1}^m P(f_k^j|c_j)$

Learning: $P(c_1) = \frac{n(c_1)+st(c_1)}{n+s}$, $P(c_i|c_1) = \frac{n(c_1,c_i)+st(c_1,c_i)}{n(c_1)+st(c_1)}$, $P(f_k^j|c_j) = \frac{n(c_j,f_k)+st(c_j,f_k^j)}{n(c_j)+st(c_j)}$

IDM: $\sum_{c_1} t(c_1) = 1 \; \sum_{c_i} t(c_1, c_i) = t(c_1), \forall c_i \; \sum_{f_k^j} t(c_j, f_k^j) = \sum_{c_1} t(c_1, c_j) = t(c_j), \forall c_j$

(Un)dominance test $\inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}',\boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'',\boldsymbol{f})} \leq 1$ is a convex optimization!

If $c'_1 = c''_1$, $\prod_{i:c'_i=\neg c''_i} \frac{n(c'_1,c'_i) \cdot \inf_{0<t_i<1} \prod_k \frac{n(c'_i,f_k)}{n(c'_i)+s(1-t_i)} \left[\frac{n(c''_i,f_k)+st_i}{n(c''_i)+st_i}\right]^{-1}}{n(c''_1,c''_i)+s} \leq 1$ (likewise if $c'_1 \neq c''_1$)

Not enough, 2^n sequences, $O(2^{2n})$ dominance tests to perform!

Given seq $\boldsymbol{c}''$, optimality $\max_{\boldsymbol{c}'} \inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}',\boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'',\boldsymbol{f})} \leq 1$ can be efficiently tested

Still not enough! 2^n optimality tests, possibly exponential output

De Bock and de Cooman: linear description of a possibly exponential output

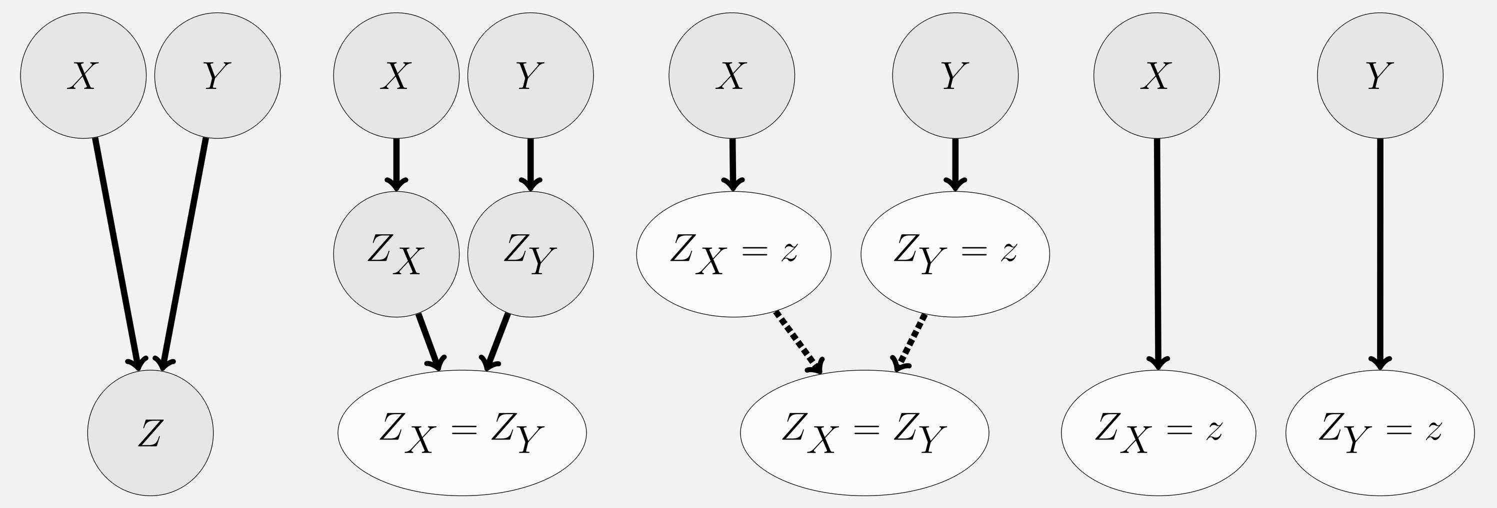
For each class label, check whether or not there are optimal sequences with the label active and with the label inactive

Task $\min_{\boldsymbol{c}'':c''_l=0/1} \max_{\boldsymbol{c}'} \inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}',\boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'',\boldsymbol{f})} \leq 1$ can be also performed efficiently

$$\min_{c''_1} \min_{c''_2,\dots,c''_q} \max_{c'_1} \max_{c'_2,\dots,c'_2} \inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}',\boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'',\boldsymbol{f})} \geq \min_{c''_1} \max_{c'_1} \min_{c''_2,\dots,c''_q} \max_{c'_2,\dots,c'_2} \inf_{\boldsymbol{t}} \frac{P_{\boldsymbol{t}}(\boldsymbol{c}',\boldsymbol{f})}{P_{\boldsymbol{t}}(\boldsymbol{c}'',\boldsymbol{f})}$$

Classifier returns a (linear-size) outer approximation of the (possibly exponential-size) set of optimal sequences. If the number of sequences is not too large, optimality of each sequence can be tested separately.

Replicating the Features



Neglecting synergetic effects of the parents to the child

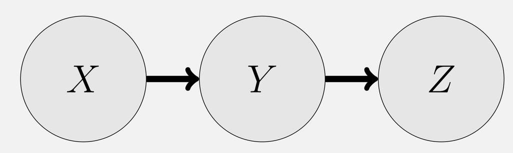
rank-one tensor decomposition (see also Ide and Cozman, 2008)

$$P(X, Y|Z) \propto P(X|Z) \cdot P(Y|Z)$$

If the child is observed, this is a duplication of Z:

$$P(X, Y|Z = z) \propto P(X|Z_X = z) \cdot P(Y|Z_Y = z)$$

IDM-based Quantification of Credal Networks



Counts s.t. $\sum_x n(x) = 1$, $\sum_y n(y, x) = n(x)$, $\sum_z n(y, z) = \sum_x n(x, z)$

IDM constraints for $X \perp\!\!\!\perp Z|Y$ (not in NCC, not in TANC)

$\sum_x t(x) = 1$, $\sum_y t(y, x) = t(x)$, $\sum_z t(y, z) = \sum_x t(x, z)$

IDM as the effect of s observation which we are totally ignorant about

Experiments

Data set	Classes	Features	Instances
Emotions	6	44/72	593
Scene	6	224/294	2407
E-mobility	10	14/18	4226
Slashdot	22	496/1079	3782

Performance described by $\%$ of instance s.t. all maximal seqs all in the same state (black points), and accuracies of the precise model when MNCC is determinate (white bars) vs. when MNCC is indeterminate (black bars)