Knowledge Intensive Learning of Cutset Networks (Supplementary Material)

Saurabh Mathur[1] **Vibhav Gogate**[1] **Sriraam Natarajan**[1]

[1]Erik Jonsson School of Engineering & Computer Science, University of Texas at Dallas, Richardson, Texas, USA

A DATASET DETAILS

Table 1 shows the number of variables, the number of qualitative influences, and the number of examples in each of the datasets.

B CUTSET NETWORK STRUCTURES

Figure 1 shows the CNs fit on the haberman data set using LearnCNet (left) and KICN (right). While the CN fit using LearnCNet selects the year of surgery as the top node (year), the CN fit using KICN selects the number of positive axillary nodes detected (Nodes) as the top node. This is in line with prior work that identifies the number of positive axillary nodes as the most important risk factor Fisher et al. [1983].

Figure 2 shows the CNs fit on the ppd data set using LearnCNet (left) and KICN (right). While the CN fit using LearnCnet selects PPD, Child care stress, and Employment status as the top variables after Age, the CN fit using KICN selects Maternity Blues and a History of Depression. Maternity Blues and a History of Depression are important risk factors for Post-partum depression [Zanardo et al., 2020, O'hara and Swain, 1996]

Figure 3 shows the CNs fit on the adni data set using LearnCNet (left) and KICN (right). While the CN fit using LearnCnet selects Education as the top variable, the CN fit using KICN selects Age as the top variable. Age is the most important risk factor in Alzheimer's Disease Knopman et al. [2021]

C ADDITIONAL RESULTS ON BENCHMARK DATA SETS

Table 2 shows the number of edges and the number of edges and the number of free parameters in the CNs fit using LearnCNet and KICN.

Data set	$\lvert\mathbf{X}\rvert$	$\lvert C\rvert$	$\lvert\mathcal{D}_{\text{train}}\rvert$	$\lvert\mathcal{D}_{\text{test}}\rvert$
cpu	7	6	104	105
ljubljana	9	6	138	139
haberman	4	3	153	153
auto	8	5	196	196
yeast	9	7	742	742
car	7	2	864	864
redwine	12	19	799	800
whitewine	12	8	2,449	2,449
abalone	9	17	2,088	2,089
cleveland	12	6	148	149
diabetes	5	4	364	365
sachs	11	6	100	100
asia	8	5	100	100
earthquake	5	3	100	100
survey	6	5	100	100
ppd	11	4	86	87
adni	7	3	168	168
numom2b-a	7	6	3,292	3,292
numom2b-b	8	7	1,828	1,829

Table 1: The number of datapoints in training ($|\mathcal{D}_{\text{train}}|$) and test sets ($|\mathcal{D}_{\text{test}}|$), number of variables ($|\mathbf{X}|$) and the number of qualitative influences ($|C|$) for all the data sets used in evaluation.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

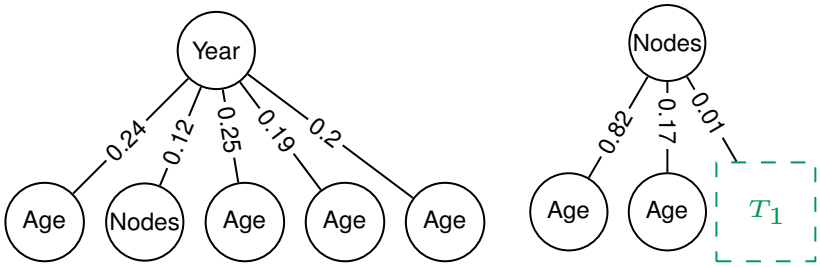

Figure 1: First two levels of cutset networks fit on a sample of the haberman dataset using LearnCNet (left) and KICN (right)

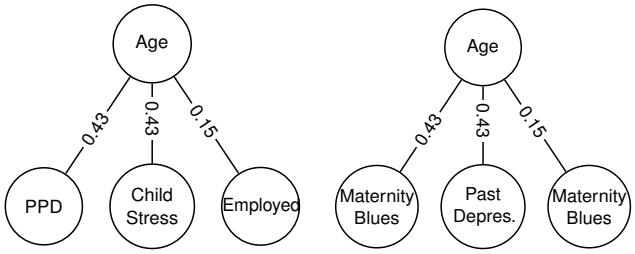

Figure 2: First two levels of cutset networks fit on a sample of the ppd dataset using LearnCNet (left) and KICN (right)

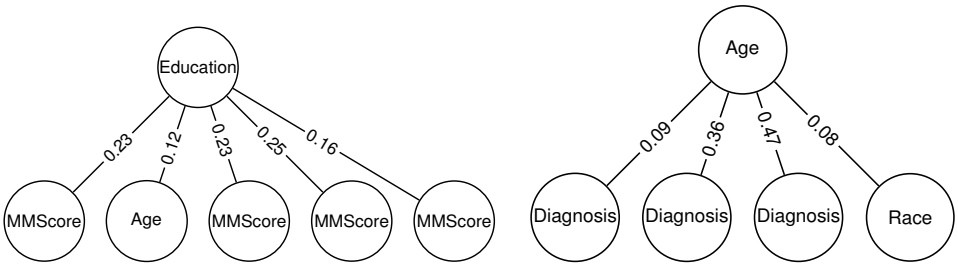

Figure 3: First two levels of cutset networks fit on a sample of the adni dataset using LearnCNet (left) and KICN (right)

	Edge count		Parameter count	
Data set	**LearnCNet**	**KICN**	**LearnCNet**	**KICN**
cpu	75.2	71.2	276.8	266.8
ljubljana	155.7	171.9	441.5	439.5
haberman	30.7	24.4	134.8	109.8
auto	173.8	205.6	724.4	735.6
car	178.7	187.6	1,115.7	1,070.7
abalone	720.0	354.4	2,914.8	1,360.4
redwine	635.4	674.8	1,006.1	978.4
whitewine	801.5	933.1	1,172.5	1,240.7
yeast	556.8	541.6	1,808.3	1,694.3
cleveland	265.30	277.70	687.00	673.70
diabetes	51.30	45.60	270.20	266.00
sachs	193.4	208.6	888.0	843.2
asia	47.4	49.4	73.3	72.9
earthquake	11.0	11.0	18.4	17.7
survey	27.7	25.9	63.8	62.8

Table 2: The number of edges and the number of free parameters for cutset networks fit using LearnCNet and KICN on UCI data sets with 30% noise (rows 1–11) and data sampled from Bayesian Networks (rows 12–15) The counts are averaged over 10 bootstrap samples.

References

Bernard Fisher, Madeline Bauer, D Lawrence Wickerham, Carol K Redmond, Edwin R Fisher, Anatolio B Cruz, Roger Foster, Bernard Gardner, Harvey Lerner, Richard Margolese, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. an nsabp update. *Cancer*, 52(9):1551–1557, 1983.

David S Knopman, Helene Amieva, Ronald C Petersen, Gäel Chételat, David M Holtzman, Bradley T Hyman, Ralph A Nixon, and David T Jones. Alzheimer disease. *Nature reviews Disease primers*, 7(1):33, 2021.

Michael W O'hara and Annette M Swain. Rates and risk of postpartum depression—a meta-analysis. *International review of psychiatry*, 8(1):37–54, 1996.

Vincenzo Zanardo, Francesca Volpe, Federico de Luca, Lara Giliberti, Arturo Giustardi, Matteo Parotto, Gianluca Straface, and Gino Soldera. Maternity blues: a risk factor for anhedonia, anxiety, and depression components of edinburgh postnatal depression scale. *The Journal of Maternal-Fetal & Neonatal Medicine*, 33(23):3962–3968, 2020.