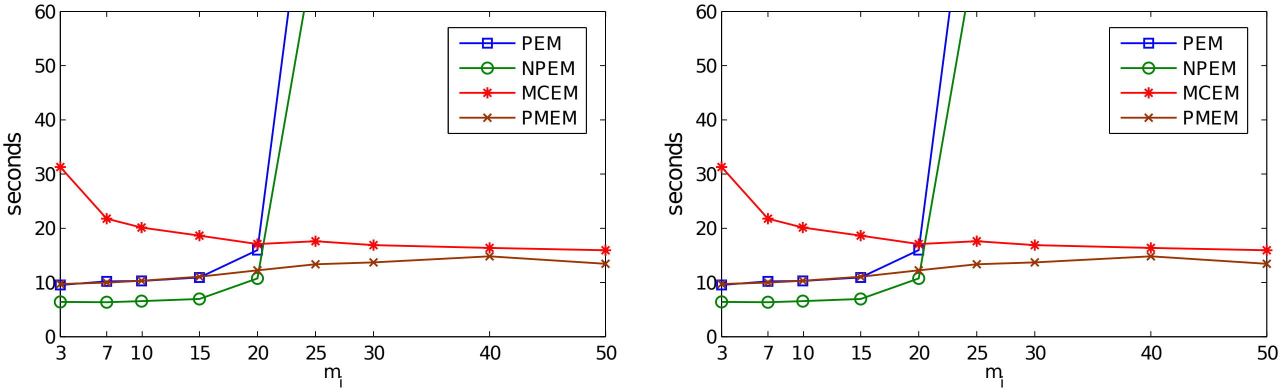
Extended experimental setting of: Learning Bayesian Network Classifiers from Label Proportions

Jerónimo Hernández, Jose A. Lozano, and Iñaki Inza

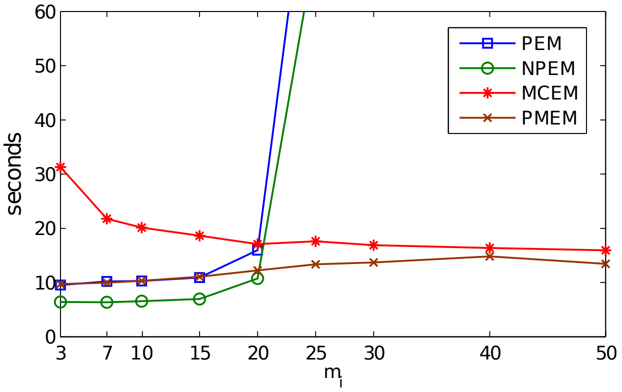
References

[1] D. R. Musicant, J. M. Christensen, J. F. Olson, Supervised learning by training on aggregate outputs, in: Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007), 2007, pp. 252–261.



(a) The proposed algorithm learning NB classifiers.

(b) The proposed algorithm learning TAN classifiers.



(c) The proposed algorithm learning 2DB classifiers.

Figure 1: Computational time needed by the four versions of the proposed algorithm to learn NB, TAN and 2DB classifiers in a 10×5 fold CV over 30 datasets (sampled from TAN models, with 31 binary variables —including the class variable— and 1000 examples). The bag size (m_i) is varied to simulate different experimental conditions. All the tests have been performed in an Intel Core i5 (2,3 GHz) with 4GB of main memory.

MLH		0.0	0.25	0.5	0.75	m_i
PEM	NB	96.34 ± 0.51	96.45 ± 0.46	96.31 ± 0.34	96.31 ± 0.40	3
	TAN	96.11 ± 0.56	96.29 ± 0.36	96.29 ± 0.22	96.47 ± 0.39	
	KDB	96.24 ± 0.41	96.14 ± 0.42	95.87 ± 0.41	95.99 ± 0.41	
NPEM	NB	96.58 ± 0.36	96.42 ± 0.30	96.48 ± 0.38	96.15 ± 0.46	
	TAN	96.25 ± 0.32	96.22 ± 0.41	96.38 ± 0.40	96.29 ± 0.46	
	KDB	95.91 ± 0.44	95.77 ± 0.32	96.19 ± 0.36	96.14 ± 0.42	
MCEM	NB	96.48 ± 0.41	96.55 ± 0.36	96.45 ± 0.38	96.27 ± 0.87	
	TAN	96.45 ± 0.28	96.49 ± 0.40	96.41 ± 0.35	96.18 ± 0.58	
	KDB	96.25 ± 0.34	95.68 ± 0.57	95.71 ± 0.41	95.84 ± 0.48	
PMEM	NB	96.62 ± 0.39	96.61 ± 0.32	96.32 ± 0.48	96.27 ± 0.40	
	TAN	95.97 ± 0.58	95.98 ± 0.41	96.24 ± 0.31	96.37 ± 0.48	
	KDB	96.01 ± 0.22	96.08 ± 0.45	95.94 ± 0.52	95.82 ± 0.39	
DT		92.69 ± 0.91	85.08 ± 1.26	71.40 ± 1.02	67.41 ± 0.65	
PEM	NB	96.48 ± 0.46	96.37 ± 0.43	96.38 ± 0.47	96.45 ± 0.41	7
	TAN	96.41 ± 0.31	96.32 ± 0.37	96.47 ± 0.44	96.37 ± 0.42	
	KDB	96.32 ± 0.59	96.24 ± 0.50	95.84 ± 0.60	96.11 ± 0.35	
NPEM	NB	96.45 ± 0.32	96.65 ± 0.19	96.38 ± 0.40	96.39 ± 0.48	
	TAN	96.17 ± 0.42	96.38 ± 0.29	96.07 ± 0.57	96.41 ± 0.56	
	KDB	96.07 ± 0.55	95.98 ± 0.65	96.15 ± 0.23	95.79 ± 0.66	
MCEM	NB	96.52 ± 0.44	96.29 ± 0.41	96.44 ± 0.36	96.38 ± 0.41	
	TAN	96.27 ± 0.45	96.37 ± 0.40	96.44 ± 0.42	96.17 ± 0.64	
	KDB	96.15 ± 0.49	95.81 ± 0.59	96.05 ± 0.50	95.82 ± 0.46	
PMEM	NB	95.97 ± 0.72	95.99 ± 0.41	96.57 ± 0.25	96.39 ± 0.43	
	TAN	96.27 ± 0.30	96.31 ± 0.28	96.41 ± 0.51	96.21 ± 0.52	
	KDB	96.12 ± 0.64	96.15 ± 0.49	95.69 ± 0.61	95.88 ± 0.50	
DT		91.86 ± 0.81	81.93 ± 1.95	70.73 ± 1.10	67.71 ± 0.78	
PEM	NB	96.41 ± 0.65	96.52 ± 0.35	96.45 ± 0.25	96.57 ± 0.37	15
	TAN	96.25 ± 0.34	96.05 ± 0.91	96.19 ± 0.28	96.31 ± 0.38	
	KDB	96.29 ± 0.32	95.84 ± 0.59	95.81 ± 0.46	96.02 ± 0.47	
NPEM	NB	96.39 ± 0.46	96.47 ± 0.32	96.17 ± 0.50	96.55 ± 0.29	
	TAN	96.42 ± 0.45	95.92 ± 0.57	96.21 ± 0.37	96.54 ± 0.48	
	KDB	96.17 ± 0.44	95.75 ± 0.55	95.77 ± 0.52	95.87 ± 0.44	
MCEM	NB	96.45 ± 0.45	95.88 ± 0.86	96.72 ± 0.36	96.28 ± 0.37	
	TAN	96.35 ± 0.47	96.39 ± 0.42	96.48 ± 0.29	96.21 ± 0.42	
	KDB	95.91 ± 0.35	96.05 ± 0.40	95.85 ± 0.48	96.01 ± 0.42	
PMEM	NB	96.05 ± 0.43	96.29 ± 0.73	96.55 ± 0.37	96.31 ± 0.32	
	TAN	96.28 ± 0.52	96.42 ± 0.37	96.28 ± 0.44	96.11 ± 0.52	
	KDB	95.68 ± 0.78	96.07 ± 0.29	95.89 ± 0.33	95.95 ± 0.42	
DT		90.36 ± 0.65	80.87 ± 1.20	71.32 ± 1.13	67.75 ± 0.75	

Table 1: Comparing our methods with Musicant et al. proposals [1]. Breast Cancer Wisconsin dataset, evaluated for increasing bag size ($m_i = \{3, 7, 15\}$) and MLH entropy (MLH $= \{0.0, 0.25, 0.5, 0.75\}$) using a 10×5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and the four versions of our method learning different Bayesian classifiers.

MLH		0.0	0.25	0.5	0.75	m_i
MCEM	NB	96.49 ± 0.44	96.61 ± 0.41	96.54 ± 0.35	96.14 ± 0.89	30
	TAN	96.35 ± 0.48	96.41 ± 0.58	96.58 ± 0.43	96.19 ± 0.33	
	KDB	96.29 ± 0.43	95.88 ± 0.45	95.94 ± 0.40	95.38 ± 0.52	
PMEM	NB	96.37 ± 0.42	96.37 ± 0.47	96.58 ± 0.35	96.51 ± 0.36	
	TAN	96.21 ± 0.31	96.19 ± 0.42	96.38 ± 0.34	96.09 ± 0.47	
	KDB	96.12 ± 0.23	95.89 ± 0.59	95.81 ± 0.55	95.92 ± 0.44	
DT		90.39 ± 0.73	81.63 ± 1.27	70.26 ± 1.78	67.83 ± 0.70	
MCEM	NB	96.35 ± 0.32	96.38 ± 0.31	96.34 ± 0.43	96.09 ± 0.77	50
	TAN	96.05 ± 0.58	96.01 ± 0.41	96.12 ± 0.39	96.01 ± 0.53	
	KDB	96.08 ± 0.47	96.09 ± 0.47	96.02 ± 0.58	95.31 ± 0.52	
PMEM	NB	96.61 ± 0.33	96.48 ± 0.41	96.29 ± 0.58	96.34 ± 0.44	
	TAN	96.14 ± 0.47	96.12 ± 0.49	95.92 ± 0.55	95.49 ± 0.44	
	KDB	95.88 ± 0.28	96.14 ± 0.34	95.29 ± 0.47	93.39 ± 0.95	
DT		86.62 ± 1.13	80.82 ± 1.70	71.14 ± 1.11	67.54 ± 0.45	

Table 2: Comparing our methods with Musicant et al. proposals [1]. Breast Cancer Wisconsin dataset, evaluated for large bag sizes ($m_i = \{30, 50\}$) and MLH entropy (MLH $= \{0.0, 0.25, 0.5, 0.75\}$) using a 10×5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and two specific versions of our method learning different Bayesian classifiers.

MLH		0.0	0.25	0.5	0.75	m_i
PEM	NB	81.20 ± 3.82	81.82 ± 3.46	78.69 ± 4.24	81.08 ± 3.77	3
	TAN	84.53 ± 2.39	85.33 ± 1.87	84.64 ± 3.25	82.68 ± 2.77	
	KDB	88.92 ± 2.28	89.57 ± 2.40	90.03 ± 2.80	89.86 ± 2.38	
NPEM	NB	85.19 ± 2.40	82.96 ± 3.06	78.55 ± 3.95	77.52 ± 4.90	
	TAN	84.76 ± 2.95	84.53 ± 2.60	84.53 ± 1.87	83.68 ± 2.86	
	KDB	89.63 ± 1.44	90.43 ± 0.98	89.34 ± 1.89	89.69 ± 2.46	
MCEM	NB	82.48 ± 3.75	83.62 ± 2.45	81.17 ± 5.48	77.49 ± 6.26	
	TAN	84.13 ± 3.12	83.59 ± 2.50	82.76 ± 3.07	83.13 ± 2.34	
	KDB	90.20 ± 1.13	90.01 ± 1.63	88.49 ± 2.10	88.86 ± 1.46	
PMEM	NB	84.07 ± 2.14	83.33 ± 2.69	78.97 ± 2.22	75.87 ± 5.42	
	TAN	84.53 ± 2.25	85.56 ± 1.42	83.56 ± 2.93	81.82 ± 2.53	
	KDB	89.09 ± 2.47	90.28 ± 1.23	89.29 ± 2.43	87.38 ± 2.37	
DT		85.33 ± 1.57	83.48 ± 1.07	72.74 ± 2.04	68.07 ± 3.55	
PEM	NB	83.70 ± 2.21	82.39 ± 3.08	79.83 ± 5.58	75.56 ± 2.23	7
	TAN	85.41 ± 3.23	83.25 ± 1.70	84.19 ± 1.74	82.99 ± 1.97	
	KDB	90.60 ± 1.68	89.12 ± 1.99	88.52 ± 2.96	87.72 ± 3.41	
NPEM	NB	85.07 ± 2.84	81.45 ± 3.46	78.21 ± 5.02	73.22 ± 3.65	
	TAN	85.44 ± 2.81	85.90 ± 2.14	84.53 ± 2.21	82.68 ± 2.82	
	KDB	88.18 ± 2.67	88.55 ± 3.52	89.40 ± 0.89	87.78 ± 3.30	
MCEM	NB	82.96 ± 2.55	80.51 ± 4.32	79.26 ± 4.43	73.48 ± 2.62	
	TAN	82.54 ± 2.43	82.17 ± 2.16	81.77 ± 2.74	82.74 ± 2.40	
	KDB	90.46 ± 1.90	88.55 ± 2.11	89.20 ± 2.87	88.38 ± 3.29	
PMEM	NB	83.59 ± 3.47	82.17 ± 3.08	79.91 ± 3.10	73.93 ± 7.35	
	TAN	83.16 ± 3.37	85.19 ± 1.82	82.45 ± 2.64	82.93 ± 2.25	
	KDB	88.75 ± 2.94	88.92 ± 1.75	88.43 ± 2.28	88.75 ± 1.80	
DT		83.65 ± 1.85	76.15 ± 2.52	67.54 ± 2.26	59.36 ± 2.53	
PEM	NB	82.11 ± 4.37	78.77 ± 4.42	81.68 ± 2.89	74.81 ± 4.96	15
	TAN	85.10 ± 2.69	87.18 ± 2.26	83.05 ± 2.72	84.27 ± 2.68	
	KDB	89.26 ± 2.40	89.89 ± 1.28	89.57 ± 1.67	87.35 ± 2.60	
NPEM	NB	80.80 ± 3.55	82.96 ± 3.60	78.92 ± 6.65	76.32 ± 4.97	
	TAN	84.39 ± 2.54	84.81 ± 1.97	83.48 ± 2.33	83.08 ± 4.67	
	KDB	89.69 ± 1.87	89.04 ± 2.03	87.09 ± 2.94	85.61 ± 3.50	
MCEM	NB	81.54 ± 3.85	81.37 ± 2.22	76.67 ± 5.17	75.58 ± 3.42	
	TAN	85.33 ± 1.76	83.02 ± 3.72	83.42 ± 2.33	83.48 ± 2.76	
	KDB	89.40 ± 1.87	88.97 ± 1.99	88.60 ± 1.76	87.81 ± 2.64	
PMEM	NB	83.93 ± 4.01	81.51 ± 5.31	81.08 ± 3.46	76.52 ± 7.06	
	TAN	84.36 ± 2.08	83.65 ± 1.86	83.22 ± 1.76	83.25 ± 2.08	
	KDB	88.58 ± 3.35	89.43 ± 1.73	90.09 ± 2.04	86.47 ± 2.82	
DT		81.23 ± 2.14	74.22 ± 1.63	64.69 ± 1.66	57.87 ± 1.91	

Table 3: Comparing our methods with Musicant et al. proposals [1]. Ionosphere dataset, evaluated for increasing bag size ($m_i = \{3, 7, 15\}$) and MLH entropy ($\text{MLH} = \{0.0, 0.25, 0.5, 0.75\}$) using a 10×5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and the four versions of our method learning different Bayesian classifiers.

MLH		0.0	0.25	0.5	0.75	m_i
MCEM	NB	82.85 ± 2.79	79.72 ± 4.78	77.04 ± 3.79	72.48 ± 3.40	30
	TAN	85.33 ± 2.14	83.96 ± 1.51	83.19 ± 2.56	83.19 ± 2.68	
	KDB	88.83 ± 2.87	90.28 ± 2.19	89.06 ± 1.41	86.58 ± 2.54	
PMEM	NB	80.28 ± 3.62	81.17 ± 3.73	78.15 ± 2.08	71.99 ± 2.50	
	TAN	84.13 ± 3.11	85.13 ± 2.64	84.25 ± 2.13	82.82 ± 3.05	
	KDB	90.26 ± 1.46	88.92 ± 2.30	88.35 ± 3.73	85.38 ± 3.40	
DT		81.87 ± 3.27	71.85 ± 4.20	63.09 ± 2.47	57.18 ± 1.35	
MCEM	NB	82.71 ± 1.65	82.76 ± 2.15	76.55 ± 5.62	75.24 ± 2.94	50
	TAN	84.90 ± 1.85	84.47 ± 1.93	84.76 ± 2.79	82.39 ± 2.52	
	KDB	88.06 ± 1.48	89.97 ± 0.52	87.75 ± 2.41	85.50 ± 2.99	
PMEM	NB	84.19 ± 4.68	81.79 ± 2.43	78.46 ± 4.29	73.82 ± 3.41	
	TAN	85.21 ± 1.33	83.93 ± 2.18	84.05 ± 2.79	80.94 ± 2.93	
	KDB	88.83 ± 1.21	89.03 ± 1.66	89.57 ± 2.37	86.47 ± 2.21	
DT		82.05 ± 1.43	72.79 ± 2.51	61.94 ± 3.53	56.05 ± 1.87	

Table 4: Comparing our methods with Musicant et al. proposals [1]. Ionosphere dataset, evaluated for large bag sizes ($m_i = \{30, 50\}$) and MLH entropy (MLH $= \{0.0, 0.25, 0.5, 0.75\}$) using a 10×5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and two specific versions of our method learning different Bayesian classifiers.

MLH		0.0	0.25	0.5	0.75	m_i
PEM	NB	99.75 ± 0.08	99.73 ± 0.05	99.73 ± 0.08	99.73 ± 0.12	3
	TAN	99.64 ± 0.21	99.75 ± 0.15	99.75 ± 0.08	99.67 ± 0.16	
	KDB	99.48 ± 0.36	99.23 ± 0.24	99.37 ± 0.25	99.54 ± 0.35	
NPEM	NB	99.70 ± 0.08	99.73 ± 0.05	99.73 ± 0.06	99.75 ± 0.08	
	TAN	99.62 ± 0.22	99.70 ± 0.08	99.67 ± 0.16	99.26 ± 0.87	
	KDB	99.34 ± 0.30	99.32 ± 0.54	99.48 ± 0.31	99.32 ± 0.33	
MCEM	NB	99.73 ± 0.04	99.73 ± 0.02	99.73 ± 0.04	99.73 ± 0.04	
	TAN	99.70 ± 0.15	99.73 ± 0.09	99.73 ± 0.12	99.67 ± 0.12	
	KDB	99.29 ± 0.37	99.40 ± 0.34	99.21 ± 0.26	99.40 ± 0.24	
PMEM	NB	99.70 ± 0.08	99.73 ± 0.12	99.73 ± 0.06	99.70 ± 0.08	
	TAN	99.71 ± 0.23	99.73 ± 0.12	99.75 ± 0.15	99.73 ± 0.17	
	KDB	99.59 ± 0.22	99.37 ± 0.42	99.34 ± 0.39	99.48 ± 0.26	
DT		96.91 ± 0.59	88.22 ± 1.19	75.85 ± 1.49	71.17 ± 0.88	
PEM	NB	99.73 ± 0.09	99.73 ± 0.12	99.73 ± 0.09	99.70 ± 0.08	7
	TAN	99.54 ± 0.25	99.75 ± 0.15	99.73 ± 0.12	99.62 ± 0.13	
	KDB	99.48 ± 0.31	99.21 ± 0.31	99.21 ± 0.51	99.18 ± 0.32	
NPEM	NB	99.73 ± 0.18	99.73 ± 0.20	99.70 ± 0.15	99.73 ± 0.19	
	TAN	99.56 ± 0.33	99.75 ± 0.19	99.64 ± 0.13	99.73 ± 0.17	
	KDB	99.48 ± 0.26	99.41 ± 0.34	99.43 ± 0.48	99.37 ± 0.39	
MCEM	NB	99.73 ± 0.05	99.75 ± 0.08	99.75 ± 0.08	99.73 ± 0.07	
	TAN	99.72 ± 0.23	99.70 ± 0.15	99.71 ± 0.08	99.70 ± 0.15	
	KDB	99.54 ± 0.21	99.29 ± 0.46	99.23 ± 0.42	99.26 ± 0.41	
PMEM	NB	99.73 ± 0.05	99.71 ± 0.08	99.67 ± 0.11	99.73 ± 0.09	
	TAN	99.64 ± 0.30	99.64 ± 0.17	99.62 ± 0.35	99.64 ± 0.13	
	KDB	99.42 ± 0.32	99.29 ± 0.41	99.21 ± 0.31	99.40 ± 0.27	
DT		96.31 ± 0.74	86.78 ± 1.92	75.77 ± 1.63	71.58 ± 0.81	
PEM	NB	99.73 ± 0.10	99.73 ± 0.12	99.73 ± 0.12	99.73 ± 0.12	15
	TAN	99.72 ± 0.15	99.73 ± 0.17	99.75 ± 0.08	99.71 ± 0.15	
	KDB	99.37 ± 0.31	99.37 ± 0.35	99.51 ± 0.24	99.51 ± 0.38	
NPEM	NB	99.73 ± 0.12	99.73 ± 0.04	99.73 ± 0.05	99.73 ± 0.05	
	TAN	99.75 ± 0.08	99.72 ± 0.08	99.54 ± 0.27	99.71 ± 0.19	
	KDB	99.32 ± 0.44	99.37 ± 0.37	99.32 ± 0.39	99.23 ± 0.44	
MCEM	NB	99.73 ± 0.07	99.71 ± 0.08	99.75 ± 0.08	99.70 ± 0.08	
	TAN	99.75 ± 0.08	99.67 ± 0.11	99.67 ± 0.16	99.62 ± 0.18	
	KDB	99.23 ± 0.36	99.40 ± 0.49	99.23 ± 0.36	99.37 ± 0.30	
PMEM	NB	99.73 ± 0.12	99.67 ± 0.11	99.73 ± 0.06	99.67 ± 0.11	
	TAN	99.62 ± 0.28	99.59 ± 0.25	99.64 ± 0.13	99.70 ± 0.08	
	KDB	99.34 ± 0.33	99.29 ± 0.39	99.29 ± 0.25	99.40 ± 0.32	
DT		94.29 ± 0.93	87.30 ± 1.10	75.44 ± 0.89	71.81 ± 1.04	

Table 5: Comparing our methods with Musicant et al. proposals [1]. Dermatology dataset, evaluated for increasing bag size ($m_i = \{3, 7, 15\}$) and MLH entropy (MLH $= \{0.0, 0.25, 0.5, 0.75\}$) using a 10×5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and the four versions of our method learning different Bayesian classifiers.

MLH		0.0	0.25	0.5	0.75	m_i
MCEM	NB	99.73 ± 0.01	99.73 ± 0.01	99.71 ± 0.08	99.71 ± 0.08	30
	TAN	99.67 ± 0.24	99.70 ± 0.15	99.73 ± 0.17	99.67 ± 0.11	
	KDB	99.29 ± 0.46	99.43 ± 0.24	99.23 ± 0.47	99.04 ± 0.48	
PMEM	NB	99.73 ± 0.04	99.71 ± 0.08	99.72 ± 0.08	99.64 ± 0.13	
	TAN	99.73 ± 0.02	99.75 ± 0.08	99.67 ± 0.24	99.70 ± 0.08	
	KDB	99.32 ± 0.22	99.18 ± 0.35	99.37 ± 0.42	99.02 ± 0.48	
DT		91.17 ± 1.34	87.76 ± 2.33	76.72 ± 0.87	72.32 ± 0.89	
MCEM	NB	99.67 ± 0.11	99.73 ± 0.12	99.73 ± 0.12	99.59 ± 0.14	50
	TAN	99.64 ± 0.13	99.67 ± 0.11	99.64 ± 0.13	99.62 ± 0.22	
	KDB	99.56 ± 0.33	99.37 ± 0.21	99.04 ± 0.60	98.55 ± 0.47	
PMEM	NB	99.73 ± 0.09	99.71 ± 0.08	99.67 ± 0.11	99.54 ± 0.17	
	TAN	99.67 ± 0.11	99.70 ± 0.08	99.59 ± 0.25	99.41 ± 0.23	
	KDB	99.48 ± 0.33	99.23 ± 0.38	98.25 ± 0.62	95.57 ± 1.64	
DT		95.71 ± 0.91	87.49 ± 1.79	74.29 ± 1.22	71.39 ± 0.95	

Table 6: Comparing our methods with Musicant et al. proposals [1]. Dermatology dataset, evaluated for large bag sizes ($m_i = \{30, 50\}$) and MLH entropy (MLH $= \{0.0, 0.25, 0.5, 0.75\}$) using a 10 × 5 fold CV. The result is the accuracy and associated standard deviation, for their best proposal (DT) and two specific versions of our method learning different Bayesian classifiers.