

Online Resource 1

Task-Conditional Accuracy–Support Trade-offs of Sparse Attention Normalizers in Compact Classifiers

Machine Learning

Özkan Canay

Department of Information Systems and Technologies, Faculty of Computer and Information Sciences, Sakarya University, Sakarya, Türkiye
canay@sakarya.edu.tr

Caption. All matched-seed softmax-reference comparisons with Holm correction applied separately to the 24-comparison main family and six-comparison KMNIST family.

Family	Dataset	Method	Δ acc.	95% CI	Δ F1	Raw p	Holm p
Main	CIFAR-10	Sparsemax	0.0118	[-0.0018, 0.0253]	0.0155	0.0813	0.5693
Main	CIFAR-10	Entmax-1.5	0.0134	[0.0032, 0.0236]	0.0174	0.0154	0.1232
Main	CIFAR-10	HAE	0.0148	[0.0058, 0.0239]	0.0179	0.0048	0.0480
Main	CIFAR-10	Top-k 0.125	0.0222	[0.0109, 0.0336]	0.0287	0.0016	0.0209
Main	CIFAR-10	Top-k 0.25	0.0263	[0.0161, 0.0365]	0.0319	0.0002	0.0051
Main	CIFAR-10	Top-k 0.50	0.0235	[0.0127, 0.0343]	0.0250	0.0008	0.0146
Main	Fashion-MNIST	Sparsemax	0.0222	[0.0163, 0.0281]	0.0262	1.33×10^{-5}	0.0003
Main	Fashion-MNIST	Entmax-1.5	0.0167	[0.0086, 0.0247]	0.0195	0.0011	0.0157
Main	Fashion-MNIST	HAE	0.0158	[0.0065, 0.0251]	0.0185	0.0038	0.0421
Main	Fashion-MNIST	Top-k 0.125	0.0260	[0.0158, 0.0362]	0.0307	0.0003	0.0055
Main	Fashion-MNIST	Top-k 0.25	0.0229	[0.0108, 0.0350]	0.0259	0.0020	0.0243
Main	Fashion-MNIST	Top-k 0.50	0.0144	[0.0050, 0.0238]	0.0171	0.0072	0.0650
Main	20 Newsgroups	Sparsemax	0.0219	[0.0118, 0.0319]	0.0243	0.0008	0.0146
Main	20 Newsgroups	Entmax-1.5	0.0449	[0.0380, 0.0517]	0.0461	1.31×10^{-7}	3.01×10^{-6}
Main	20 Newsgroups	HAE	0.0463	[0.0393, 0.0532]	0.0478	1.06×10^{-7}	2.55×10^{-6}
Main	20 Newsgroups	Top-k 0.125	0.0245	[0.0132, 0.0359]	0.0267	0.0009	0.0146
Main	20 Newsgroups	Top-k 0.25	0.0257	[0.0151, 0.0362]	0.0273	0.0004	0.0073
Main	20 Newsgroups	Top-k 0.50	0.0245	[0.0132, 0.0359]	0.0262	0.0009	0.0146
Main	Synthetic	Sparsemax	-0.0003	[-0.0046, 0.0041]	0.0008	0.8992	1.0000
Main	Synthetic	Entmax-1.5	0.0003	[-0.0050, 0.0056]	0.0006	0.9013	1.0000
Main	Synthetic	HAE	-0.0003	[-0.0060, 0.0054]	-0.0005	0.9078	1.0000
Main	Synthetic	Top-k 0.125	-0.0033	[-0.0103, 0.0037]	-0.0036	0.3149	1.0000
Main	Synthetic	Top-k 0.25	-0.0001	[-0.0072, 0.0071]	-0.0011	0.9877	1.0000
Main	Synthetic	Top-k 0.50	0.0020	[-0.0035, 0.0075]	0.0003	0.4326	1.0000
KMNIST	KMNIST	Sparsemax	0.0123	[0.0037, 0.0210]	0.0117	0.0105	0.0381
KMNIST	KMNIST	Entmax-1.5	0.0134	[0.0042, 0.0226]	0.0130	0.0092	0.0381
KMNIST	KMNIST	HAE	0.0143	[0.0042, 0.0244]	0.0137	0.0106	0.0381
KMNIST	KMNIST	Top-k 0.125	0.0126	[0.0010, 0.0242]	0.0126	0.0358	0.0381
KMNIST	KMNIST	Top-k 0.25	0.0186	[0.0063, 0.0308]	0.0179	0.0076	0.0381
KMNIST	KMNIST	Top-k 0.50	0.0152	[0.0068, 0.0236]	0.0147	0.0027	0.0162

Notes. Differences are method minus softmax over ten matched seeds. Confidence intervals are unadjusted two-sided 95% t intervals with $df = 9$; Holm adjustment is applied separately within the two reporting families stated in the manuscript.