

dataset	#data	#feature	Description	Task	Metric
Higgs Boson	10M	28	Dense	Classification	AUC
KDD10	19M	29M	Sparse	Binary classification	AUC
Credit	7M	107	Dense & Sparse	Regression	MAE
CIFAR-10 & CIFAR-100	60k	32*32	Image	Classification	Accuracy
dataset	#Train	#Test	#Words	#Classes	Average Length
20NG	11,314	7,532	42K	20	221
R8	5,485	2,189	7K	8	65
R52	6,532	2,568	9K	52	70
Ohsumed	3,357	4,043	14K	23	135
MR	7,108	3,554	18K	2	20
SST	8,544	3,311	18K	5	18

Table 1: Statistics of the datasets used in the experiments.

1 A Additional Details