Table 1: Performance on multiple large datasets with Joint Learning

	AmazonCat-13K	Wiki10K-31K	Delicious-200K	WikiLSHTC-325K	Wikipedia-500K
Prec@1	93.05	86.82	47.70	62.15	62.27
Prec@2	86.56	80.44	43.67	48.37	49.60
Prec@3	79.18	74.30	41.22	39.58	41.43
Prec@4	72.06	68.61	39.37	33.52	35.68
Prec@5	64.54	63.68	37.98	29.10	31.42
nDCG@1	93.05	86.82	47.70	62.15	62.27
nDCG@2	89.60	81.89	44.63	57.00	55.23
nDCG@3	87.72	77.22	42.75	55.20	52.11
nDCG@4	86.35	72.92	41.34	54.77	50.59
nDCG@5	85.92	69.13	40.27	54.84	49.88