Dataset	# Query	# Train	# Dim
MNIST	10,000	60,000	784
NEWS20	2,000	18,000	1,355,191
NYTIMES	5,000	100,000	102,660
RCV1	5,000	100,000	47,236
URL	5,000	90,000	3,231,958
WEBSPAM	5,000	100,000	16,609,143