An Experiment

- data:
  - instances x uniform from $\{-1, +1\}^{10,000}$
  - label y = majority vote of three coordinates
  - weak classifier = single coordinate (or its negation)
  - training set size $m = 1000$
- algorithms (all provably minimize exponential loss):
  - standard AdaBoost
  - gradient descent on exponential loss
  - AdaBoost, but in which weak classifiers chosen at random
- results:

exp. loss	% test error [# rounds]		
	stand. AdaB.	grad. desc.	random AdaB.
10^{-10}	0.0 [94]	40.7 [5]	44.0 [24,464]
10^{-20}	0.0 [190]	40.8 [9]	41.6 [47,534]
10^{-40}	0.0 [382]	40.8 [21]	40.9 [94,479]
10^{-100}	0.0 [956]	40.8 [70]	40.3 [234,654]