Boosting and Noise

[Long & Servedio]

- can construct data source that "breaks" AdaBoost with even tiny amount of noise (say, 1%)
  - Bayes optimal error $= 1\%$
    (obtainable by classifier of same form as AdaBoost)
  - AdaBoost provably has error $\geq 50\%$
- holds even if:
  - given unlimited training data
  - use any method for minimizing exponential loss
    (holds for most other convex losses as well)
- shows:
  - consistency result can fail badly if weak classifiers "not rich enough"
  - boosting susceptible to noise
- on "real-world" datasets, AdaBoost often works anyway
- various theoretical algorithms proposed for handling noise (e.g., [Kalai & Servedio], [Long & Servedio])