More Technically...

- with high probability, $\forall \theta > 0$:

$$\text{generalization error} \leq \hat{\Pr}[\text{margin} \leq \theta] + \tilde{O}\left(\frac{\sqrt{d/m}}{\theta}\right)$$

$(\hat{\Pr}[\,] = \text{empirical probability})$

- bound depends on
    - $m = \#$ training examples
    - $d = $ "complexity" of weak classifiers
    - entire distribution of margins of training examples
- $\hat{\Pr}[\text{margin} \leq \theta] \to 0$ exponentially fast (in T)
  if $\epsilon_t < \frac{1}{2} - \theta$ ($\forall t$)
    - so: if weak learning assumption holds, then all examples
      will quickly have "large" margins