

# 1、机器可学习性分析



► 测验：根据霍夫丁不等式

$$P(|E_{in}(h) - E_{out}(h)| > \varepsilon) \leq 2M \exp(-2\varepsilon^2 N) = \delta$$

可由给定的容忍误差 $\varepsilon$ 和对坏样本的容忍误差 $\delta$ 来确定需要收集多大的数据集 $N$ 来满足该要求。给定 $\varepsilon = 0.1$ ,  $\delta = 0.05$ ,  $M = 100$ ,那么需要多大的数据集。

A、 215

B、 415

C、 615

D、 815



$$N = \frac{1}{2\varepsilon^2} \ln \frac{2M}{\delta}$$