Statistical Learning Theory: Error Bounds and VC-Dimension

CS6780 – Advanced Machine Learning
Spring 2015
Thorsten Joachims
Cornell University

Reading:
Schoelkopf/Smola Chapter 5 (remainder)

Vapnik Chervonenkis Dimension

• Definition: The VC-Dimension of H is equal to the maximum number d of examples that can be split into two sets in all $2^d$ ways using functions from H (shattering).

Generalization Error Bound: Infinite H, Non-Zero Error

• Setting
  – Sample of n labeled instances $S$
  – Learning Algorithm $L$ using a hypothesis space $H$ with $VCDim(H)=d$
  – $L$ returns hypothesis $\hat{h}=L(S)$ with lowest training error
• Given hypothesis space $H$ with $VCDim(H)$ equal to $d$ and an i.i.d. sample $S$ of size $n$, with probability (1-δ) it holds that

\[
F_{\infty}(\hat{h}_S) \leq F_{\infty}(h_{\infty}) \cdot \sqrt{d} \frac{n}{\delta} \left(\frac{d}{n} + 1\right) \cdot \ln \left(\frac{2}{\delta}\right)
\]

VC Dimension of Hyperplanes

• Theorem: The VC Dimension of unbiased hyperplanes over N features is N.
• Theorem: The VC Dimension of biased hyperplanes over N features is N+1.