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

$$Err_P(h_{\mathcal{L}(S)}) \leq Err_S(h_{\mathcal{L}(S)}) + \sqrt{\frac{d\left(\ln\left(\frac{2n}{d}\right)+1\right)-\ln\left(\frac{\delta}{4}\right)}{n}}$$

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.