Foundations of Data Science & Machine Learning

Summary — Week 07
Devansh Singh Rathore
111701011

B.Tech. in Computer Science & Engineering
Indian Institute of Technology Palakkad

May 11, 2021

Abstract

x x

1 Uniform Convergence

→ We proved that H is PAC learnable iff VC-dimension of H is finite.
→ What more can one ask for?

1. Can we tolerate a non-zero in-sample error?
2. Can we use combination of simple (small VC-dimension) classifier?
3. Anything between d and d+1 for VC-dimension?
4. Boosting

1.1 Non-Zero In-Sample Error

Goal:

$$P_{S\sim D^n}[\exists h \in H : |E_D(h,f) - E_S(h,f)| > \epsilon] \leq \delta$$
$$\equiv P_{S\sim D^n}[\cup_{h\in H} C_h] \leq \delta$$
$$(where\ C_h = |E_D(h,f) - E_S(h,f)| > \epsilon)$$

Step 1: Fix an arbitrary $h \in H$
Let $\mu = E_D[h,f]$ (a deterministic quantity)
Let $X = E_S[h,f]$ (random variable)
$= (1/n)|x \in S : h(x) \neq f(x)|$
$= (1/n)(X_1 + X_2 + ... + X_n)$
where,

$$X_i = \begin{Bmatrix} 1 & if\ h(x_i) \neq f(x_i) \\ 0 & o/w \end{Bmatrix}$$

So X is a bernoulli random variable (0-1 random variable).

$S = \{x_1, ..., x_n\}$
So X is the average of n independent bernoulli random variables.
$EX = (1/n) \sum EX_i$
$= \sum P_{x_i \sim D}[h(x_i) \neq f(x_i)]$
$= (1/n) \sum E_D(h, f)$
$= (1/n) \sum \mu$
$= \mu = E_D(h, f)$
Hence $C_h = |X - \mu_X| > \epsilon$

Hoeffding Bounds:
Let $X_1, X_2, ..., X_n$ be independent $\{0, 1\}$ random variables with $P(X_i = 1) = p = EX_i \ \forall i$ $(EX = (1/n) \sum EX_i = p)$. Let $X = (1/n)(X_1 + ... + X_n)$, *then*
(a) $P(X > p + \epsilon) \leq e^{-2\epsilon^2 n}$
(b) $P(X < p - \epsilon) \geq e^{-2\epsilon^2 n}$
So $P[|X - p| > \epsilon] \leq 2e^{-2\epsilon^2 n}$

Hence $P_{S \sim D^n}[C_h] \leq 2e^{-2\epsilon^2 n}$

Recall:

$$P_{S \sim D^n}[A_h] \leq (1 - \epsilon)^n \leq e^{-\epsilon n}$$

$$A_h = (E_S(h, f) = 0) \wedge (E_D(h, f) > \epsilon)$$

$$B_h = (E_S(h, f) = 0) \wedge (E_{S'}(h, f) > \epsilon/2)$$

Now,

$$C_h = |E_D(h, f) - E_S(h, f)| > \epsilon$$

$$D_h = |E_{S'}(h, f) - E_S(h, f)| > \epsilon/2$$

Claim: $P(C_h/D_h) \geq 1/2$
Proof:
Let $D'_h = |E_{S'}(h, f) - E_D(h, f)| \leq \epsilon/2$
Given C_h, $D'_h/C_h \Rightarrow D_h/C_h$
$P(D_h/C_h) \geq P(D'_h/C_h)$
$\geq P(|X - \mu_X| \leq \epsilon/2)$
$\geq 1 - 2e^{-n\epsilon^2/2}$
$\geq 1/2 \ if \ n > 4/\epsilon^2$
Hence $P(D/C) \geq 1/2 \ when \quad - (1)$

$$D = \cup_{h \in H} D_h \ and \ C = \cup_{h \in H} C_h$$