17th March

Theory of Generalization

- Training v/s Testing

e.g. practice problems vs Exam problems

- Testing

Terms: E_{in} → training error

E_{out} → ~~test error~~ ~~training error~~ test error

ϵ → ~~small~~ threshold

N → No of training instances

$$\boxed{P[|E_{in} - E_{out}| > \epsilon] \leq 2 e^{-2\epsilon^2 N}}$$

- ~~Testing~~ Training

M - No of hypothesis

$$P[|E_{in} - E_{out}| > \epsilon] \leq 2M e^{-2\epsilon^2 N}$$

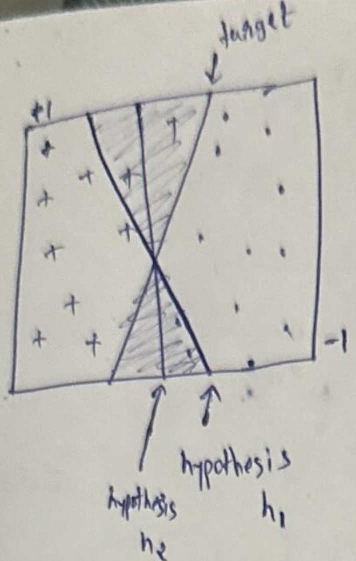
- Bad events

$B_m \rightarrow |E_{in}(h_{(m)}) - E_{out}(h_m)| > \epsilon$

Apply union bound

$$P[B_1 \text{ or } B_2 \text{ or} \ldots B_m] \leq P[B_1] + P[B_2] + \ldots P[B_m]$$

M is risky as for a perceptron M can lead to ∞



$E_{out} \rightarrow$ shaded region

$E_{in} \rightarrow$ points in shaded region

Here $|E_{in}(h_1) - E_{out}(h_1)| = |E_{in}(h_2) - E_{out}(h_2)|$

Here hypothesis are overlapping so it's meaningless to consider M as hypothesis count.

Now instead of whole hypothesis space consider data points

- **Dichotomies**

We add new term - dichotomy $\longrightarrow$ (basically subset of all data points)

Hypothesis	Dichotomies
$h: X \rightarrow \{-1, +1\}$	$h: \{x_1, x_2, \dots x_N\} \rightarrow \{-1, +1\}$ $H(x_1, x_2, \dots x_N) = \{h(x_1), h(x_2), \dots h(x_N)\}$
No. of hypothesis $\lvert h \rvert$ can be infinite	No. of dichotomies $\lvert H(x_1, x_2, \dots x_N) \rvert \le 2^N$

So we can replace No. of hypothesis with no of dichotomies

- **Growth function**

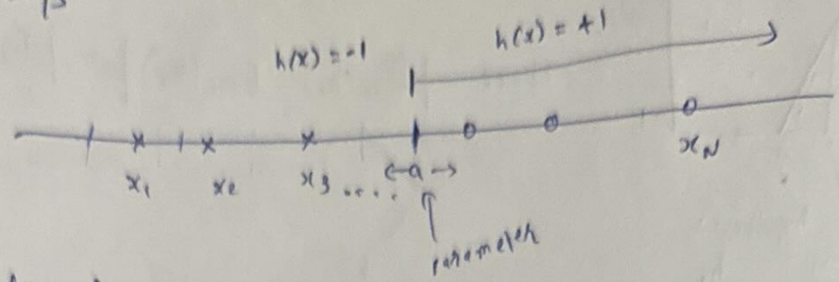
Growth function $m_H(N)$ is the maximum no. of dichotomies that can be generated by H on any N points.

$$m_H(N) = \max_{x_1, x_2 \dots x_N \in X} |H(x_1, x_2, \dots x_N)|$$

$$m_H(N) \le 2^N$$

- Value of Growth function for some cases

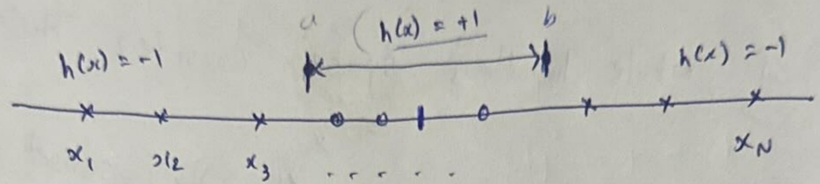
1) Positive rays



$h(x) = \text{sign}(x - a)$

$$\boxed{m_H(N) = N + 1}$$

2) Positive Intervals

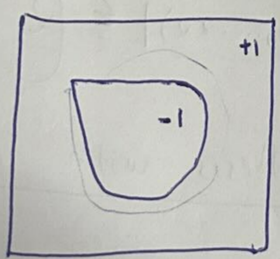


Here we have to choose two points of interval from N+1.

$$m_H(N) = \binom{N+1}{2} + 1 \longrightarrow$$ for all red or no interval

$$= \frac{1}{2}N^2 + \frac{1}{2}N + 1$$

3) Convex sets



$m_H(N) = 2^N$

- In $P[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2M e^{-2\epsilon^2 N}$, replace M with the growth function $m_H(N)$

- Break Point

→ If no data set of size k can be shattered by H, then k is said to be the break point for H.

→ If k is break point then,

$$m_H(k) < 2^k$$

→ For two dimension perceptron, k = 4 is the break point

$$m_H(4) < 2^4$$

- keep $m_H(N) = 2^N$ in the equation

$$P[|E_{in} - E_{out}| > \epsilon] \leq 2 m_H(N) e^{-2\epsilon^2 N}$$