

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/332110565>

Lecture 5: Growth function

Presentation · April 2019

CITATIONS

0

READS

1,101

1 author:



[Alaa Tharwat](#)

Fachhochschule Bielefeld

120 PUBLICATIONS 6,195 CITATIONS

SEE PROFILE

Lecture 5: Growth function

Alaa Tharwat

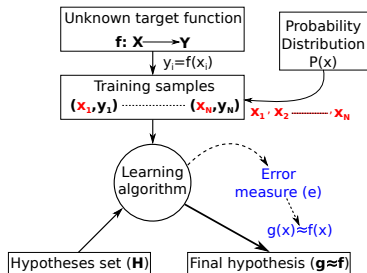
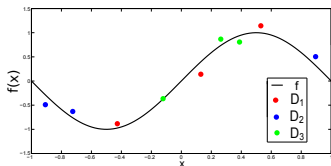
Lecture 5: Growth function

- Review of Lecture 4
- Union bound
- Illustrative examples
- Break point

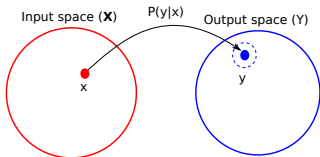
Lecture 5: Growth function

- Review of Lecture 4
- Union bound
- Illustrative examples
- Break point

- Error measure or loss function is used for evaluating a hypothesis ($E(h, f)$)
- The errors are not the same
- Noisy target \equiv deterministic target $f(\mathbf{x})$ plus noise $y - f(\mathbf{x})$
 - $P(\mathbf{x})$ is the input distribution
 - $P(y|\mathbf{x})$ is the target distribution ($y \sim P(y|\mathbf{x})$)
 - $P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$ is the mix of the two concepts
- $g \approx f \Rightarrow E_{out}(g) \approx 0$
 - 1 $E_{in}(g) \approx 0$
 - 2 $E_{out} \approx E_{in}$



| | | True Class | |
|-----------------|-----------|---------------------|---------------------|
| | | Positive (P) | Negative (N) |
| Predicted Class | True (T) | True Positive (TP) | False Positive (FP) |
| | False (F) | False Negative (FN) | True Negative (TN) |
| | | $P = TP + FN$ | $N = FP + TN$ |



Lecture 5: Growth function

- Review of Lecture 4
- **Union bound**
- Illustrative examples
- Break point

- In lecture 2:

$$\begin{aligned}
 P[|E_{in}(g) - E_{out}(g)| > \epsilon] &\leq \sum_{i=1}^M P[|E_{in}(h_i) - E_{out}(h_i)|] \\
 &\leq \sum_{i=1}^M 2e^{-2\epsilon^2 N} \\
 &\leq 2Me^{-2\epsilon^2 N}
 \end{aligned}$$

Training

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

Testing

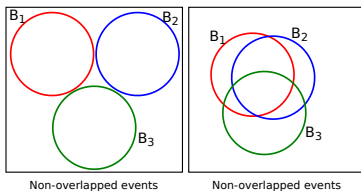
$$P[|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 2e^{-2\epsilon^2 N}$$

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

- The union¹ bound²

$$P[B_1 \text{ or } B_2 \text{ or } \dots B_M] \leq P[B_1] + P[B_2] + \dots + P[B_M]$$

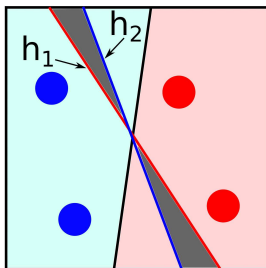
- This means that the events are not overlapped (worst case, i.e., all events are disjoint \rightarrow Extreme bad case)



¹ $P(A \text{ or } B) = P(A \cup B) \leq P(A) + P(B)$

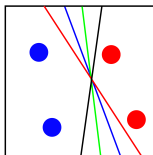
² B means bad event

- Bad events are very overlapping
- Given two hypotheses
 - ΔE_{in} ; change in in-sample error ($E_{in}(h_1) - E_{in}(h_2)$)
 - ΔE_{out} ; change in out-of-sample error, i.e., change in the area ($E_{out}(h_1) - E_{out}(h_2)$)
- The two hypotheses are overlapped, or E_{in} tracks E_{out} in h_1 approximately similar to h_2
 $\rightarrow |E_{in}(h_1) - E_{out}(h_1)| \approx |E_{in}(h_2) - E_{out}(h_2)|$



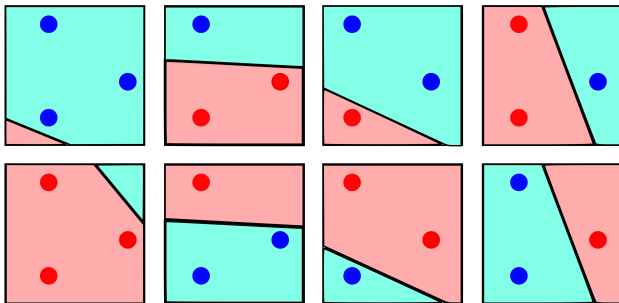
Can we replace M (worst case) with a smaller number?

- Instead of counting the number of hypotheses in the whole input space, we will focus on a finite set of input points
- This is because we cannot see all the input space, but we can estimate the results (E_{in}) of changing a hypothesis h
- Many hypotheses have the same result/output (E_{in}), because these hypotheses are overlapped
- We will consider all similar hypotheses, one hypothesis; this will lead to the so-called **dichotomy**
- Dichotomies are mini-hypotheses: $h : \{x_1, x_2, \dots, x_N\} \rightarrow \{-1, +1\}$, while the hypothesis $h : X \rightarrow \{-1, +1\}$. Hence, the dichotomies are hypotheses which are restricted to the data points
- Number of hypotheses $|H|$ is infinite, but, number of dichotomies $|H(x_1, x_2, \dots, x_N)|$ is at most 2^N (instead of M)



- Given N points, the maximum number of dichotomies is called **growth function**, and it is denoted by $m_H(N)$
- $m_H(N) = \max_{x_1, x_2, \dots, x_N \in X} |H(x_1, x_2, \dots, x_N)|$ (set of dichotomies, or, maximum number of dichotomies)
- $m_H(N) \leq 2^N$ (this is much better than $M \approx \infty$)

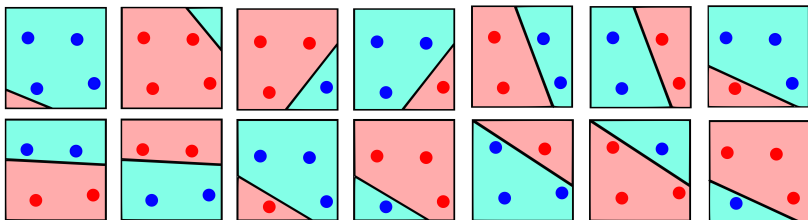
- For example, $N = 3 \Rightarrow m_H(3) = 2^3 = 8$



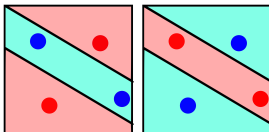
The three points can be colinear, (1) what is the growth function? (2) is the growth function changes by changing the positions of points, if yes, (3) how to calculate it?

- Answer: (1) $m_H(3) = 8$, (2) yes, (3) take the maximum

- For example, $N = 4 \Rightarrow m_H(4) = 14 \leq 2^4 = 16$



- Two hypotheses that cannot be generated in the two-dimensional space using a linear classifier

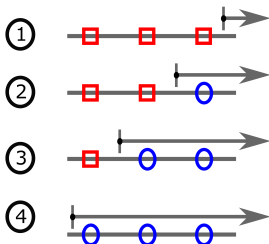
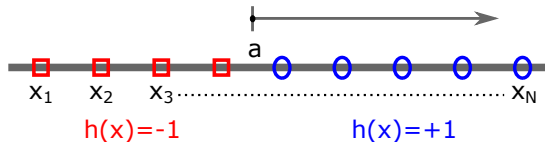


Lecture 5: Growth function

- Review of Lecture 4
- Union bound
- Illustrative examples
- Break point

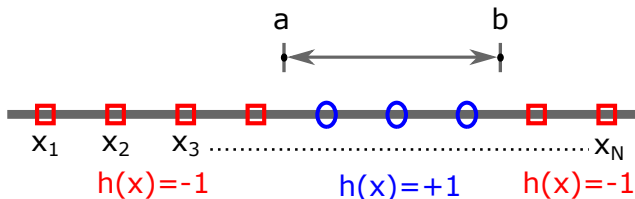
Example 1: positive rays example

- $R \rightarrow \{\pm 1\}$
- $h(x) = \text{sign}(x - a)$
- $m_H(N) = N + 1$



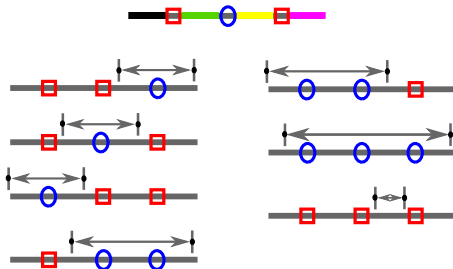
Example 2: positive intervals example

- $R \rightarrow \{\pm 1\}$
- Given N points, place intervals ends in two of $N + 1$ spots/segments
- The growth function is how to select two points from $N + 1$ segments $\left(\binom{N+1}{2}\right)$
- If all samples are blue, start and end points will be in the first and end segments. But, when all samples are red, the start and end points will be in the same segment (this is missing)
- $m_H(N) = \binom{N+1}{2} + 1 = \frac{1}{2}N^2 + \frac{1}{2}N + 1$ (Quadratic)



Example 2: positive intervals example

- The figure below shows the number of segments given three points



The growth function is

- In positive rays example: $m_H(N) = N + 1$
- In positive intervals example: $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

Replacing the growth function ($m_H(N)$) instead of M in Hoeffding's inequality is good? why?

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

- Answer: because $m_H(N)$ is polynomial (the prove in the next lecture)

Lecture 5: Growth function

- Review of Lecture 4
- Union bound
- Illustrative examples
- Break point

break point

if no dataset of size k can be shattered by H , then k is a **break point** for H

- The growth function will be $m_H(k) < 2^k$
- For example, with 2D perceptron, $k = 4$, why?
 - Answer: because the data which consists of four points cannot be shattered, so, what about bigger data?
 - Sure, cannot be shattered also

- In positive rays example ($m_H(N) = N + 1$):
 - $N = 1 \Rightarrow m_H(N) = 2 \leq 2^N = 2$
 - $N = 2 \Rightarrow m_H(N) = 3 \leq 2^N = 4$ (break point)
- In Positive intervals example ($m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$)
 - $N = 1 \Rightarrow m_H(N) = 2 \leq 2^N = 2$
 - $N = 2 \Rightarrow m_H(N) = 4 \leq 2^N = 4$
 - $N = 3 \Rightarrow m_H(N) = 7 \leq 2^N = 8$ (break point)

| | N | | | | | |
|--------------------|---|---|---|----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | ... |
| Positive rays | 2 | 3 | 4 | 5 | 2 | ... |
| Positive Intervals | 2 | 4 | 7 | 11 | 16 | ... |
| 2D Perceptron | 2 | 4 | 8 | 14 | ... | ... |

- In positive rays example: $m_H(N) = N + 1$, break point (k) = 2
- In positive intervals example: $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$, $k = 3$
- In 2D perceptron example: $k = 4$

break point

- No break point $\Rightarrow m_H(N) = 2^N$
- Any break point $\Rightarrow m_H(N)$ is **polynomial** in N

- Given three points ($N = 3$), and the break point is 2 ($k = 2$), how many rows/dichotomies we can get with these three points?
- If we have 100 points, for any choice of any three columns you cannot possibly have all possible combinations.

Break point is a very good restriction

