COMP5212: Machine Learning

Lecture 9

Minhao Cheng

From last time

Shattered

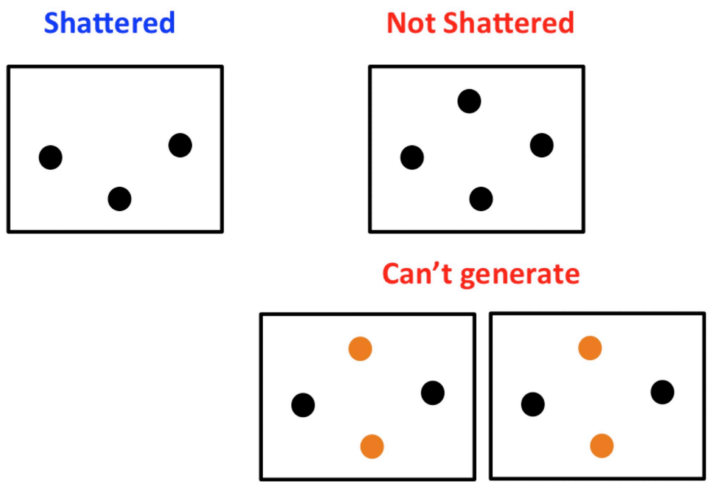
- Given a set $S = \{x^{(i)}, \ldots, x^{(d)}\}$ (no relation to the training set) of points $x^{(i)} \in \mathscr{X}$, we say that $\mathscr{H}$ shatters S if $\mathscr{H}$ can realize any labeling on S. I.e, if for any set of labels $\{y^{(i)}, \ldots, y^{(d)}\}$, there exist some $h \in \mathscr{H}$ so that $h(x^{(i)}) = y^{(i)}$ for all $i = 1,\ldots, d$

Break point of $\mathscr{H}$

- If no data set of size k can be shattered by $\mathscr{H}$, then k is a break point for $\mathscr{H}$
  - $m_{\mathscr{H}}(k) < 2^k$
- VC dimension of $\mathscr{H}$: $k - 1$ (the most points $\mathscr{H}$ can shatter)

Break point of $\mathcal{H}$

- If no data set of size k can be shattered by $\mathcal{H}$, then k is a break point for $\mathcal{H}$
  - $m_{\mathcal{H}}(k) < 2^k$
- VC dimension of $\mathcal{H}$: $k - 1$ (the most points $\mathcal{H}$ can shatter)
- For 2-D perceptron: $k = 4$, VC dimension = 3



Break point - examples

- Positive rays: $m_{\mathscr{H}}(N) = N + 1$
  - Break point $k = 2$, $d_{VC} = 1$

Break point - examples

- Positive rays: $m_{\mathscr{H}}(N) = N + 1$
  - Break point $k = 2$, $d_{VC} = 1$
- Positive intervals: $m_{\mathscr{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$
  - Break point $k = 3$, $d_{VC} = 2$

Break point - examples

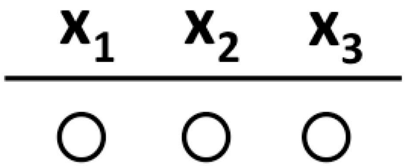
- Positive rays: $m_{\mathscr{H}}(N) = N + 1$
  - Break point $k = 2$, $d_{VC} = 1$
- Positive intervals: $m_{\mathscr{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$
  - Break point $k = 3$, $d_{VC} = 2$
- Convex set: $m_{\mathscr{H}}(N) = 2^N$
  - Break point $k = \infty$, $d_{VC} = \infty$
- Connection to # of parameters

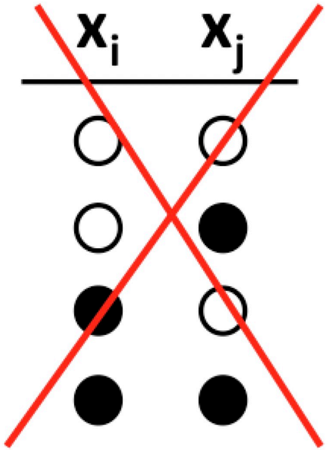
We will show

- No break point $\Rightarrow m_{\mathscr{H}}(N) = 2^N$
- Any break point $\Rightarrow m_{\mathscr{H}}(N)$ is polynomial in N

Puzzle

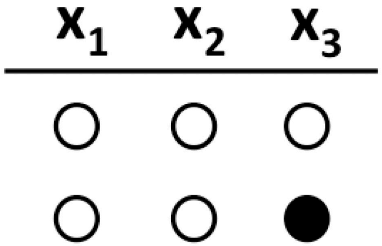
- Break point is $k = 2$

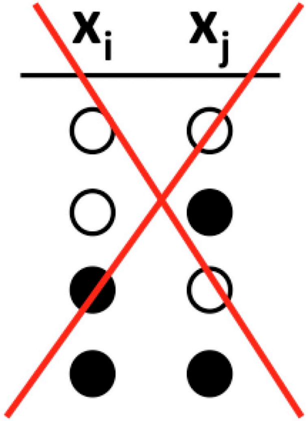




Puzzle

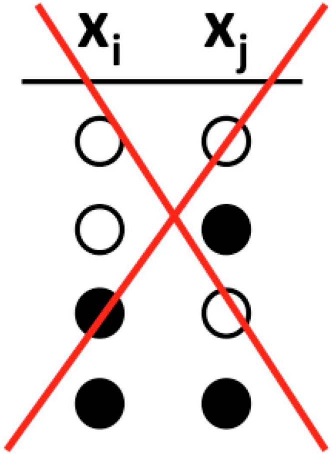
- Break point is $k = 2$

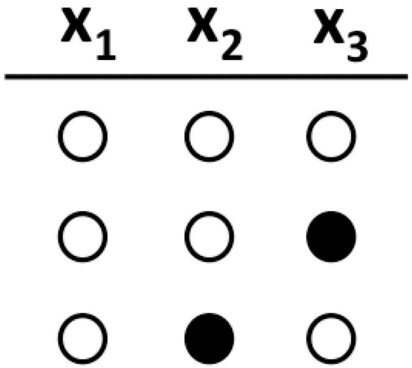




Puzzle

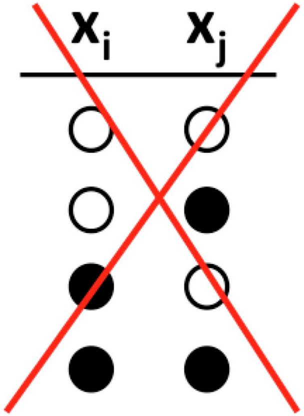
- Break point is $k = 2$

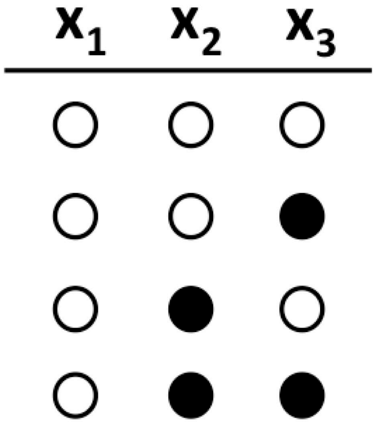




Puzzle

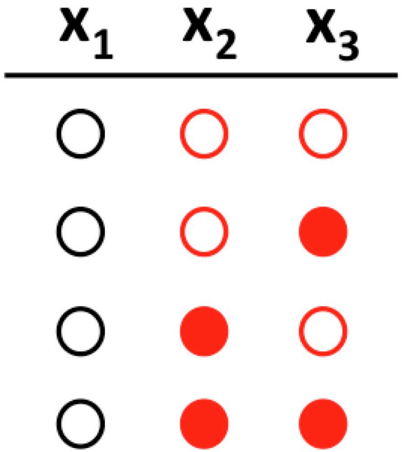
- Break point is $k = 2$

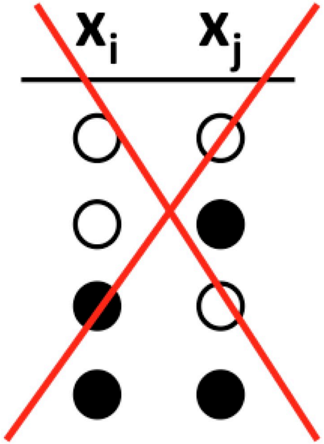




Puzzle

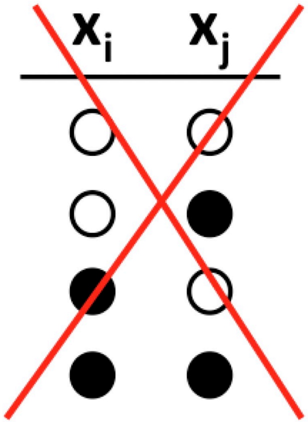
- Break point is $k = 2$

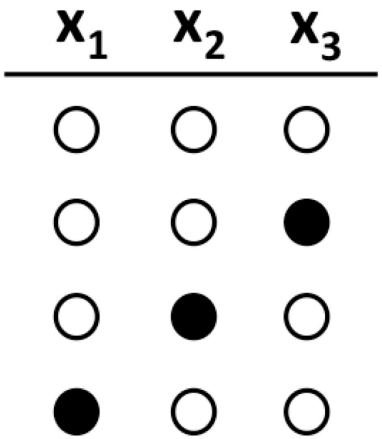




Puzzle

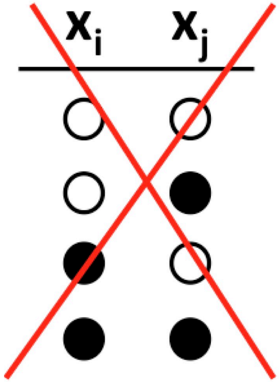
- Break point is $k = 2$

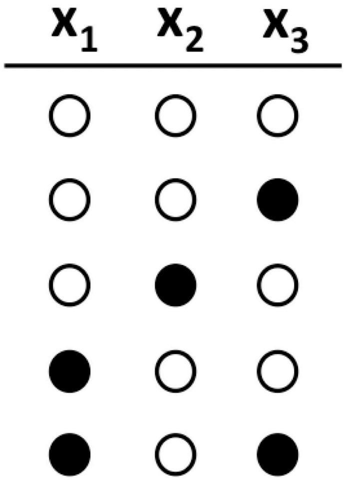




Puzzle

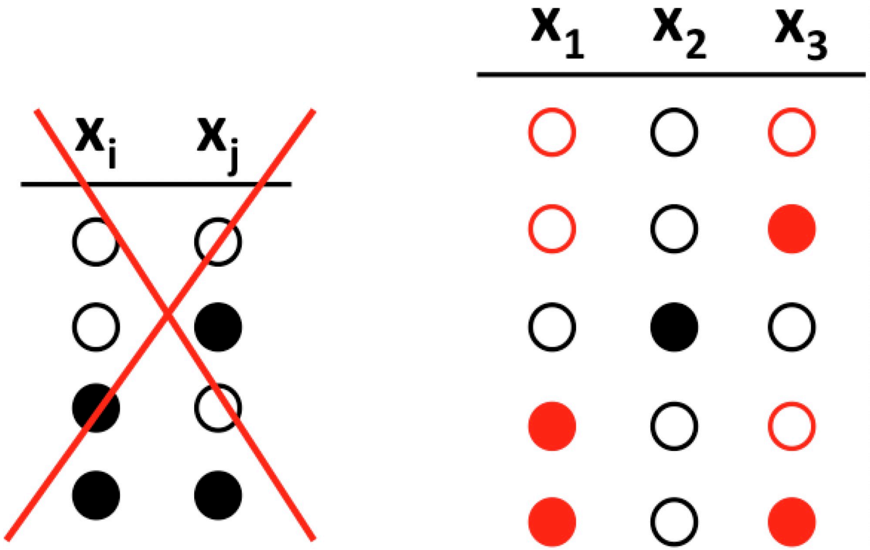
- Break point is $k = 2$





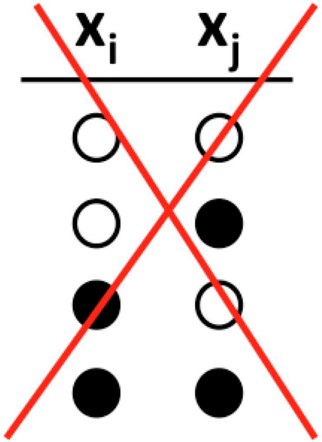
Puzzle

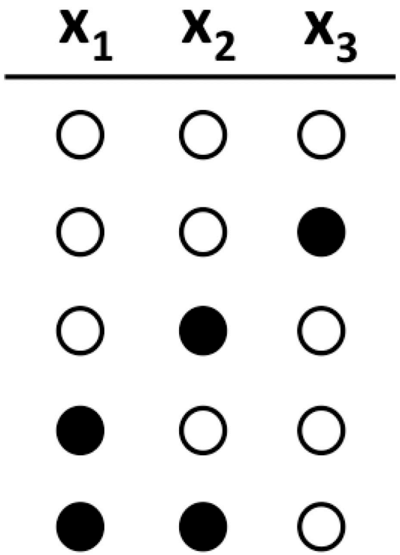
- Break point is $k = 2$



Puzzle

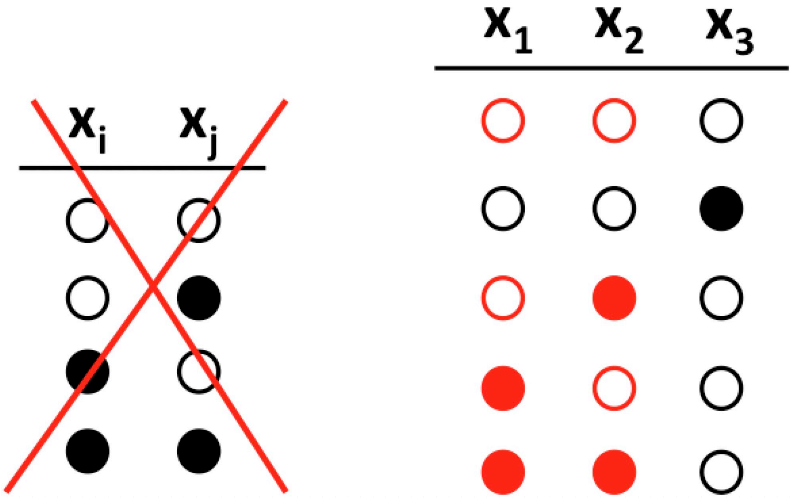
- Break point is $k = 2$





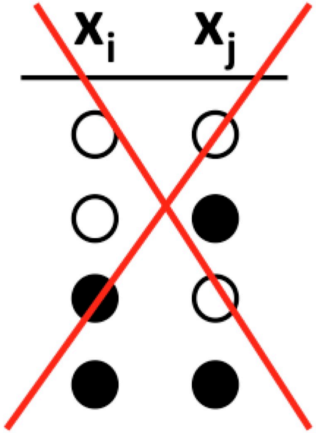
Puzzle

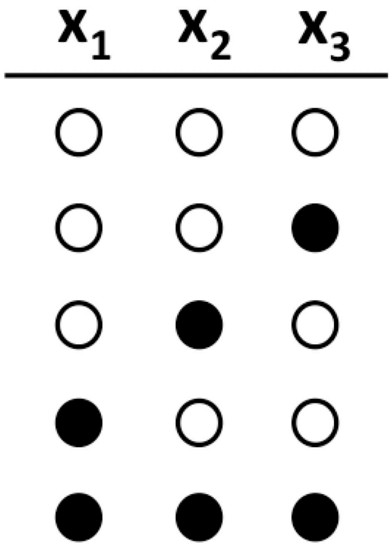
- Break point is $k = 2$



Puzzle

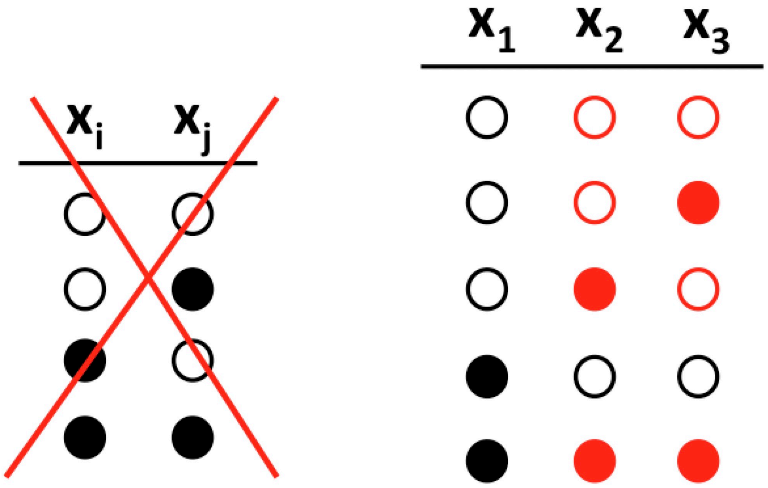
- Break point is $k = 2$





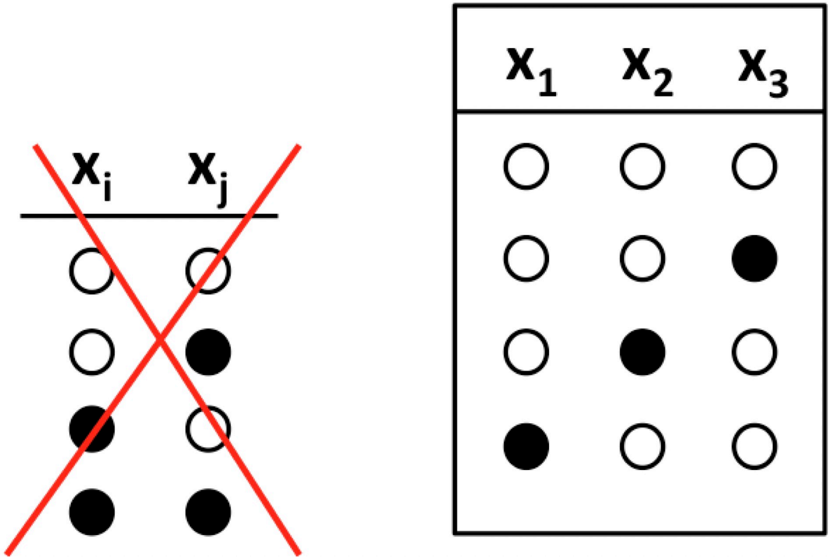
Puzzle

- Break point is $k = 2$



Puzzle

- Break point is $k = 2$



Bounding $m_{\mathscr{H}}(N)$

- Key quantity:
    - $B(N, k)$: Maximum number of dichotomies on N points, with break k

Bounding $m_{\mathscr{H}}(N)$

- Key quantity:
  - $B(N, k)$: Maximum number of dichotomies on N points, with break k
- If the hypothesis space has break point k, then
  - $m_{\mathscr{H}}(N) \leq B(N, k)$

Recursive bound on $B(N, k)$

- For any “valid” set of dichotomies, reorganize rows by
  - S_1 : pattern of $x_1, \ldots, x_{N-1}$ only appears once
  - S_2^+, S_2^- : pattern of $x_1, \ldots, x_{N-1}$ only appears twice

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\ldots$	$+1$	$+1$
			-1	$+1$	$\ldots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	-1	-1
			-1	$+1$	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	$+1$
			-1	-1	$\ldots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	$+1$
S_2			-1	-1	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	-1	-1

Recursive bound on $B(N, k)$

- Focus on $x_1, \ldots, x_{N-1}$ columns:
  - $\alpha + \beta \leq B(N-1, k)$

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
			$+1$	$+1$	$\ldots$	$+1$	$+1$
			-1	$+1$	$\ldots$	$+1$	-1
	S_1	α	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	-1	-1
			-1	$+1$	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	$+1$
			-1	-1	$\ldots$	$+1$	$+1$
	S_2^+	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	$+1$
S_2			-1	-1	$\ldots$	-1	$+1$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	$+1$	-1
	S_2^-	β	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\ldots$	$+1$	-1
			-1	-1	$\ldots$	-1	-1

Recursive bound on $B(N, k)$

- Now focus on the $S_2 = S_2^+ \cup S_2^-$:
  - $\beta \leq B(N-1, k-1)$

		# of rows	$\mathbf{x}_1$	$\mathbf{x}_2$	$\dots$	$\mathbf{x}_{N-1}$	$\mathbf{x}_N$
S_1		α	$+1$	$+1$	$\dots$	$+1$	$+1$
			-1	$+1$	$\dots$	$+1$	-1
			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	-1	-1
			-1	$+1$	$\dots$	-1	$+1$
S_2	S_2^+	β	$+1$	-1	$\dots$	$+1$	$+1$
			-1	-1	$\dots$	$+1$	$+1$
			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	$+1$	$+1$
			-1	-1	$\dots$	-1	$+1$
	S_2^-	β	$+1$	-1	$\dots$	$+1$	-1
			-1	-1	$\dots$	$+1$	-1
			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			$+1$	-1	$\dots$	$+1$	-1
			-1	-1	$\dots$	-1	-1

Recursive bound on $B(N, k)$

- $$\begin{aligned} B(N, k) &= \alpha + \beta + \beta \\ &\le B(N-1, k) + B(N-1, k-1) \end{aligned}$$
- What's the upper bound for $B(N, k)$?

Recursive bound on $B(N, k)$

- $$\begin{aligned} B(N, k) &= \alpha + \beta + \beta \\ &\le B(N-1, k) + B(N-1, k-1) \end{aligned}$$

		k					
		1	2	3	4	5	......
	1						
	2						
N	3						
	4						
	5						
	.						
	.						

Recursive bound on $B(N, k)$

- $$B(N,k) = \alpha + \beta + \beta \\ \le B(N-1,k) + B(N-1,k-1)$$

		k					
		1	2	3	4	5	……
	1	1					
	2	1					
N	3	1					
	4	1					
	5	1					
	.	.					
	.	.					

Recursive bound on $B(N, k)$

- $$B(N,k) = \alpha + \beta + \beta \leq B(N-1,k) + B(N-1,k-1)$$

		k					
		1	2	3	4	5	......
	1	1	2	2	2	2	......
	2	1					
N	3	1					
	4	1					
	5	1					
	.	.					
	.	.					

Recursive bound on $B(N, k)$

- $$B(N,k) = \alpha + \beta + \beta$$
  $$\leq B(N-1,k) + B(N-1,k-1)$$

		k					
		1	2	3	4	5	......
	1	1	2	2	2	2	......
	2	1	3				
N	3	1					
	4	1					
	5	1					
	.	.					
	.	.					

Recursive bound on $B(N, k)$

- $$\begin{aligned} B(N,k) &= \alpha + \beta + \beta \\ &\le B(N-1,k) + B(N-1,k-1) \end{aligned}$$

		k					
		1	2	3	4	5	......
	1	1	2	2	2	2	......
	2	1	3	4	4	4	......
N	3	1					
	4	1					
	5	1					
	.	.					
	.	.					

Recursive bound on $B(N, k)$

- $$\begin{aligned} B(N,k) &= \alpha + \beta + \beta \\ &\le B(N-1,k) + B(N-1,k-1) \end{aligned}$$

		k					
		1	2	3	4	5	……
	1	1	2	2	2	2	……
	2	1	3	4	4	4	……
N	3	1	4	7	8	8	……
	4	1	5	11	……		
	5	1	6	.	.		
	.	.	.	.		.	
	.	.	.				

Analytic solution for $B(N, k)$ bound

- $B(N, k)$ is upper bounded by $C(N, k)$
    - $C(N,1) = 1,\ N = 1,2,\ldots$
    - $C(1,k) = 2,\ k = 2,3,\ldots$
    - $C(N, k) = C(N-1,k) + C(N-1,k-1)$
- Theorem: $C(N, k) = \sum_{i=0}^{k=1} \binom{N}{i}$

Analytic solution for $B(N, k)$ bound

- $B(N, k)$ is upper bounded by $C(N, k)$
  - $C(N,1) = 1,\ N = 1,2,\ldots$
  - $C(1,k) = 2,\ k = 2,3,\ldots$
  - $C(N, k) = C(N-1,k) + C(N-1,k-1)$
- Sauer's Lemma: $C(N, k) = \displaystyle\sum_{i=0}^{k=1} \binom{N}{i}$
- Boundary conditions: (easy to check)

Analytic solution for $B(N,k)$ bound

- $B(N,k)$ is upper bounded by $C(N,k)$
  - $C(N,1) = 1,\ N = 1,2,\ldots$
  - $C(1,k) = 2,\ k = 2,3,\ldots$
  - $C(N,k) = C(N-1,k) + C(N-1,k-1)$
- Sauer's Theorem: $C(N,k) = \sum_{i=0}^{k-1} \binom{N}{i}$
- Boundary conditions: (easy to check)
- Induction:
  - $$\underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{select } <k \text{ from } N \text{ items}} = \underbrace{\sum_{i=0}^{k-1} \binom{N-1}{i}}_{N\text{-th item not chosen}} + \underbrace{\sum_{i=0}^{k-2} \binom{N-1}{i}}_{N\text{-th item chosen}}$$

It is polynomial!

- For a given $\mathscr{H}$, the break point k is fixed:
  - $$m_{\mathscr{H}}(N) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{Polynomial with degree } k-1}$$

It is polynomial!

- For a given $\mathscr{H}$, the break point k is fixed:
  - $$m_{\mathscr{H}}(N) \le \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{Polynomial with degree } k-1}$$
  - $\mathscr{H}$ is positive rays: (break point $k = 2$)
    - $m_{\mathscr{H}}(N) = N + 1$

It is polynomial!

- For a given $\mathscr{H}$, the break point k is fixed:
  - $$m_{\mathscr{H}}(N) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{Polynomial with degree } k-1}$$
  - $\mathscr{H}$ is 2D perceptrons: (break point $k = 4$)
    - $m_{\mathscr{H}}(N) = \,?$

It is polynomial!

- For a given $\mathcal{H}$, the break point k is fixed:
  - $$m_{\mathcal{H}}(N) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{Polynomial with degree } k-1}$$
  - $\mathcal{H}$ is 2D perceptrons: (break point $k = 4$)
    - $$m_{\mathcal{H}}(N) \leq \frac{1}{6}N^3 + \frac{5}{6}N + 1$$

Replace M by $m_{\mathscr{H}}(N)$

- Original bound:
  - $\mathbb{P}[\,\exists h \in \mathscr{H} \text{ s.t. } |E_{\text{tr}}(h) - E(h)| > \epsilon] \leq 2Me^{-2\epsilon^2 N}$
- Replace M by $m_{\mathscr{H}}(N)$
  - $\underbrace{\mathbb{P}[\,\exists h \in \mathscr{H} \text{ s.t. } |E_{\text{tr}}(h) - E(h)| > \epsilon]}_{BAD} \leq 2 \cdot 2m_{\mathscr{H}}(2N) \cdot e^{-\frac{1}{8}\epsilon^2 N}$
  - Vapnik-Chervonenkis (VC) bound

VC dimension

Definition

- The VC dimension of a hypothesis set $\mathscr{H}$, denoted by $d_{\mathsf{VC}}(\mathscr{H})$, is the largest value of N for which $m_{\mathscr{H}}(N) = 2^N$
  - "The most points $\mathscr{H}$ can shatter"

VC dimension

Definition

- The VC dimension of a hypothesis set $\mathscr{H}$, denoted by $d_{\text{VC}}(\mathscr{H})$, is the largest value of N for which $m_{\mathscr{H}}(N) = 2^N$
  - "The most points $\mathscr{H}$ can shatter"
- $N \leq d_{\text{VC}}(\mathscr{H}) \Rightarrow \mathscr{H}$ can shatter N points

VC dimension

Definition

- The VC dimension of a hypothesis set $\mathscr{H}$, denoted by $d_{\text{VC}}(\mathscr{H})$, is the largest value of N for which $m_{\mathscr{H}}(N) = 2^N$
    - "The most points $\mathscr{H}$ can shatter"
- $N \leq d_{\text{VC}}(\mathscr{H}) \Rightarrow \mathscr{H}$ can shatter N points
- $k > d_{\text{VC}}(\mathscr{H}) \Rightarrow \mathscr{H}$ cannot be shattered
- The smallest **break point** is 1 above VC-dimension

VC dimension

The growth function

- In terms of a break point k:
  - $$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$
- In terms of the VC dimension d_{VC}:
  - $$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{\text{VC}}} \binom{N}{i}$$

VC dimension

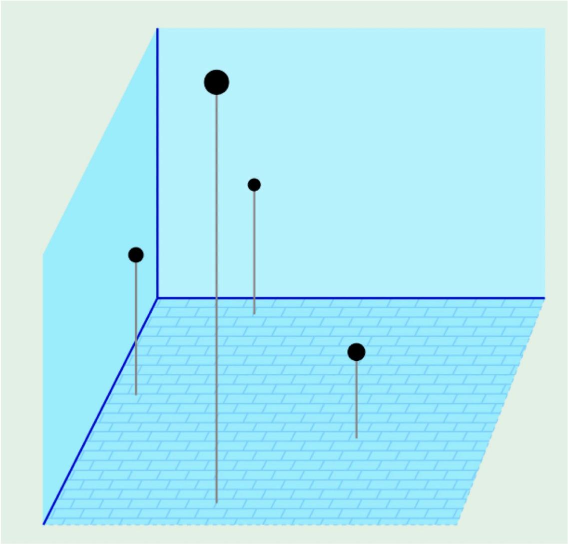
VC dimension of linear classifier

- For $d = 2$, $d_{\text{VC}} = 3$

VC dimension

VC dimension of linear classifier

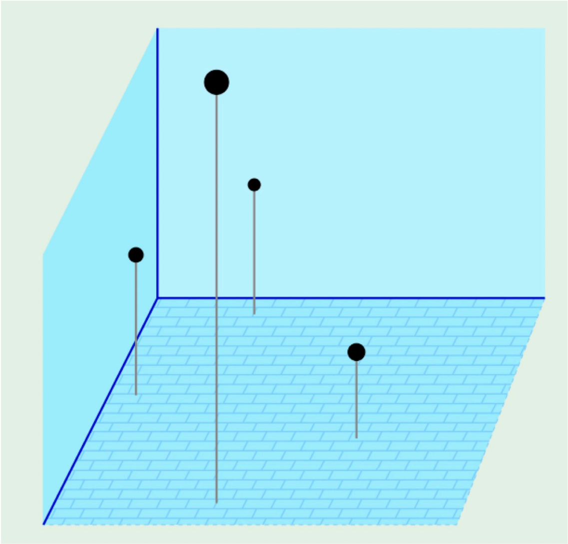
- For $d = 2$, $d_{\mathrm{VC}} = 3$
- What if $d > 2$?



VC dimension

VC dimension of linear classifier

- For $d = 2$, $d_{\text{VC}} = 3$
- What if $d > 2$?
- In general,
    - $d_{\text{VC}} = d + 1$



VC dimension

VC dimension of linear classifier

- For $d = 2$, $d_{\mathrm{VC}} = 3$
- What if $d > 2$?
- In general,
  - $d_{\mathrm{VC}} = d + 1$
- We will prove $d_{\mathrm{VC}} \geq d + 1$ and $d_{\mathrm{VC}} \leq d + 1$