

COMP5212: Machine Learning

Lecture 8

Minhao Cheng

From last time

What uniform convergence tell us?

$$P[\neg \exists h \in \mathcal{H} | E_{tr}(h) - E(h) | > \epsilon] = P[\forall h \in \mathcal{H} | E_{tr}(h) - E(h) | \leq \epsilon]$$

- $\geq 1 - 2 |\mathcal{H}| e^{-2\epsilon^2 N}$

- Given N and some δ , we have

- $|E_{tr}(h) - E(h)| \leq \sqrt{\frac{1}{2N} \log \frac{2|\mathcal{H}|}{\delta}}$

- i.e $|E_{tr}(h) - E(h)| \leq \gamma$ for all $h \in \mathcal{H}$

From last time

Shattered

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

From last time

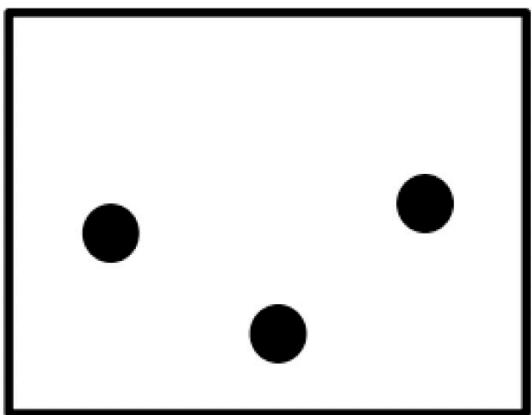
Break point of \mathcal{H}

- 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)

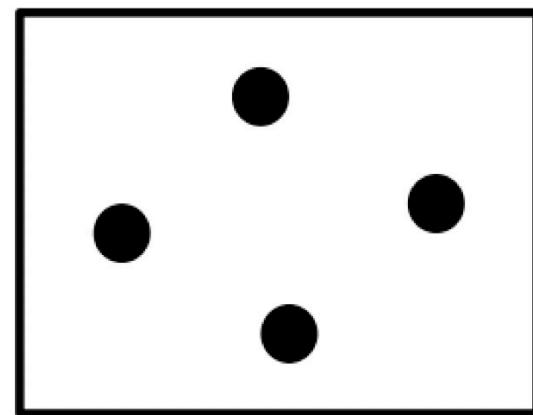
Break point of \mathcal{H}

- 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

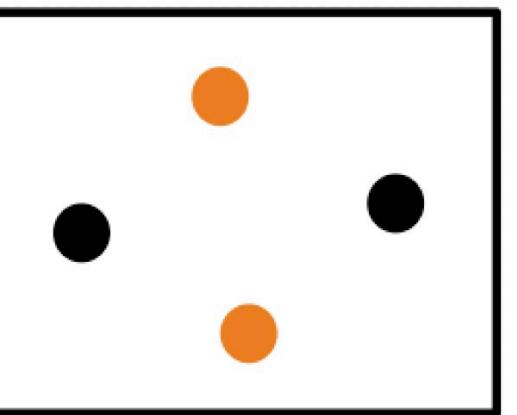
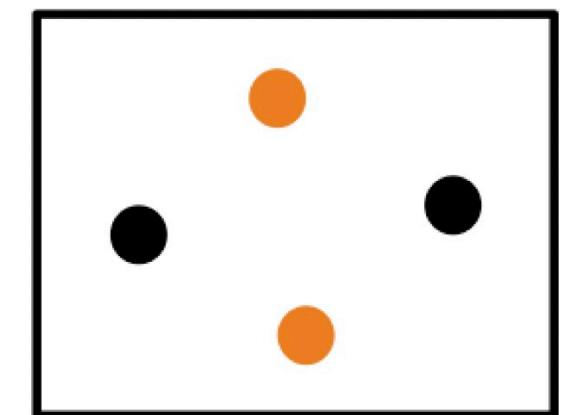
Shattered



Not Shattered



Can't generate



Break point - examples

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

Break point - examples

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

Break point - examples

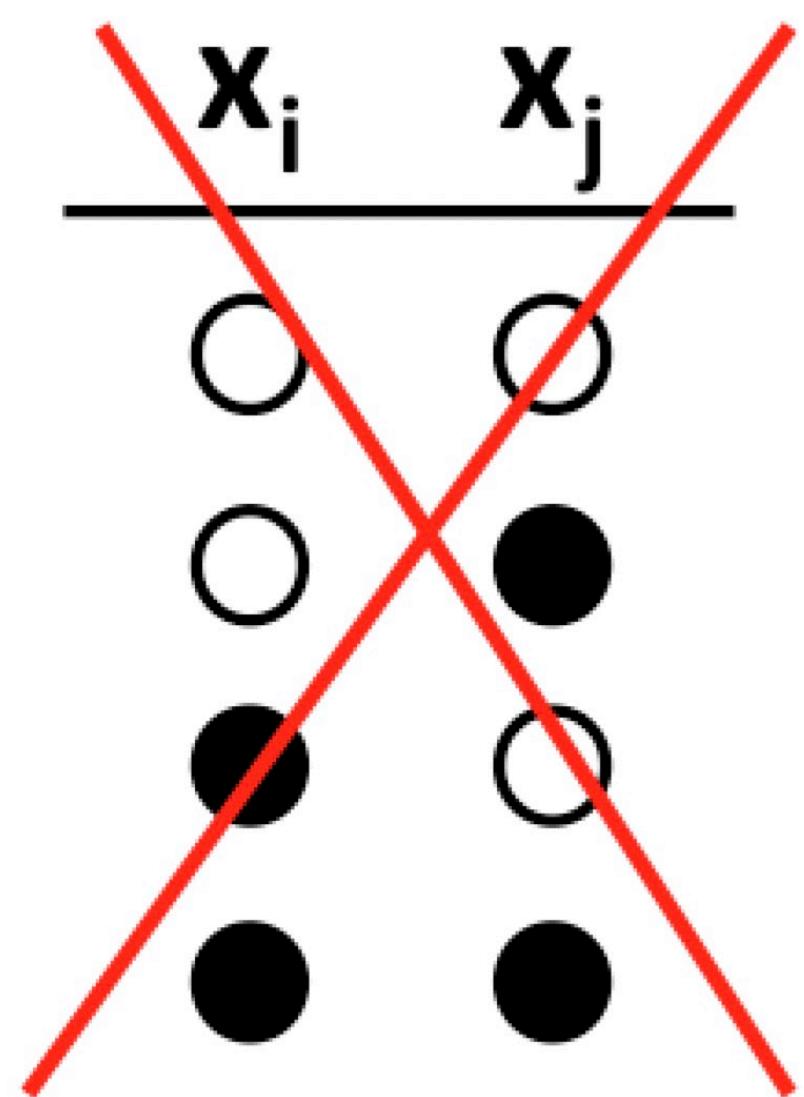
- Positive rays: $m_{\mathcal{H}}(N) = N + 1$
 - Break point $k = 2, d_{VC} = 1$
- Positive intervals: $m_{\mathcal{H}}(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$
 - Break point $k = 3, d_{VC} = 2$
- Convex set: $m_{\mathcal{H}}(N) = 2^N$
 - Break point $k = \infty, d_{VC} = \infty$
 - Connection to # of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

We will show

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

Puzzle

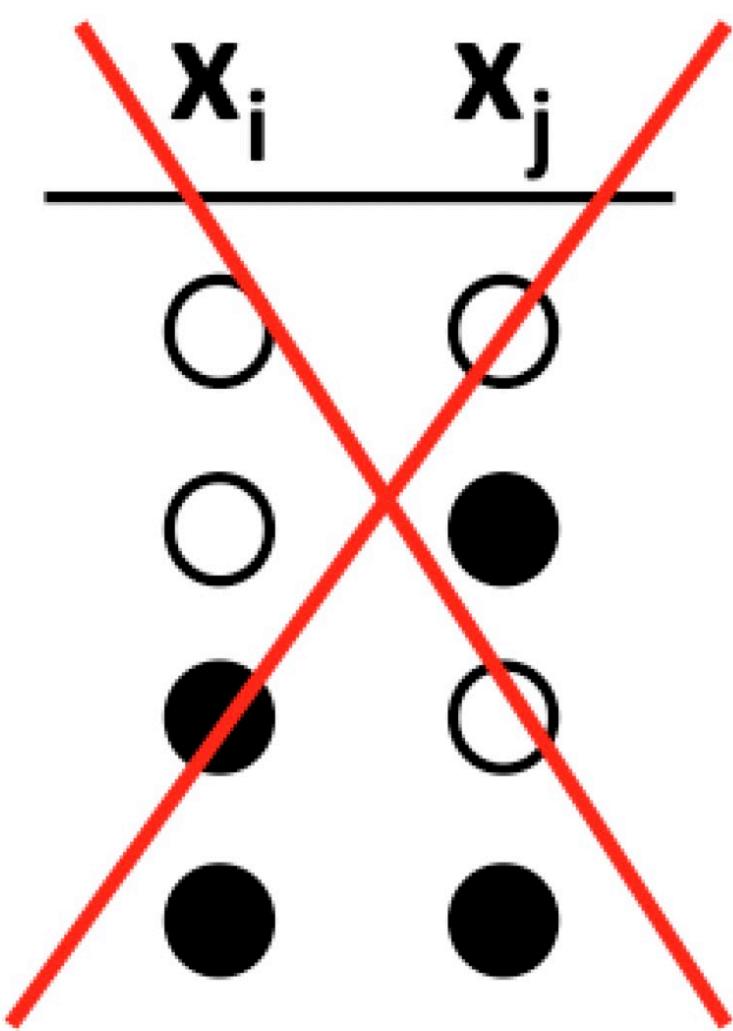
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○

Puzzle

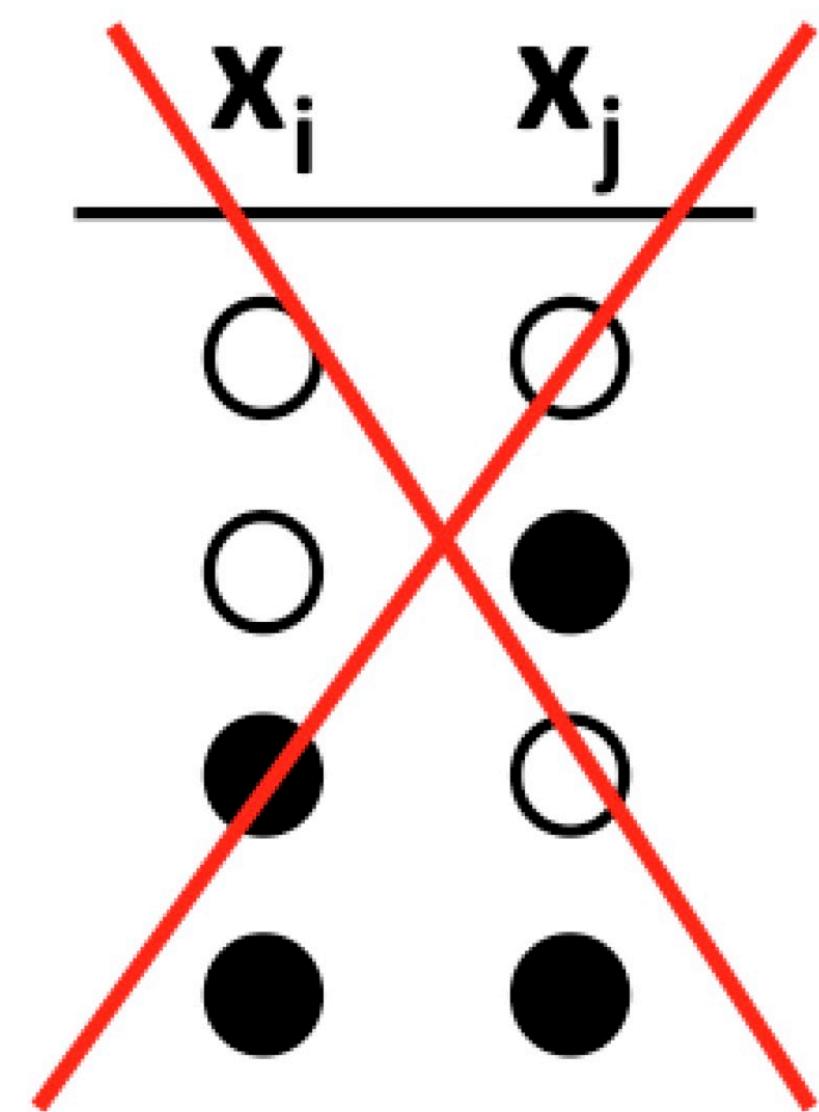
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●

Puzzle

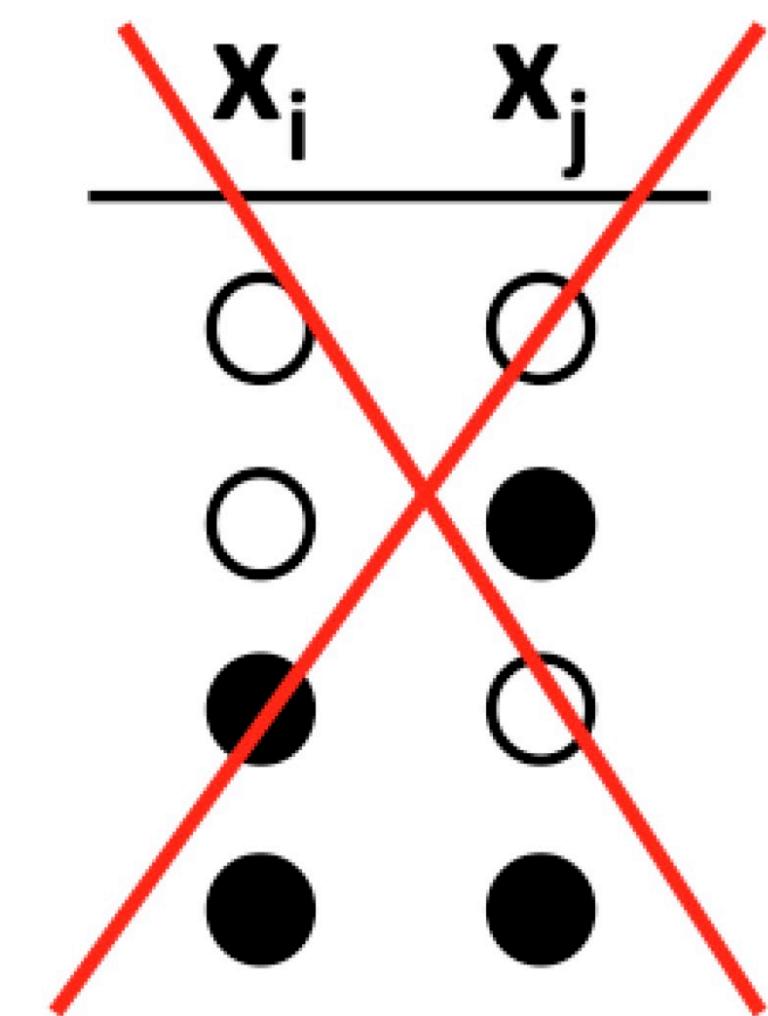
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○

Puzzle

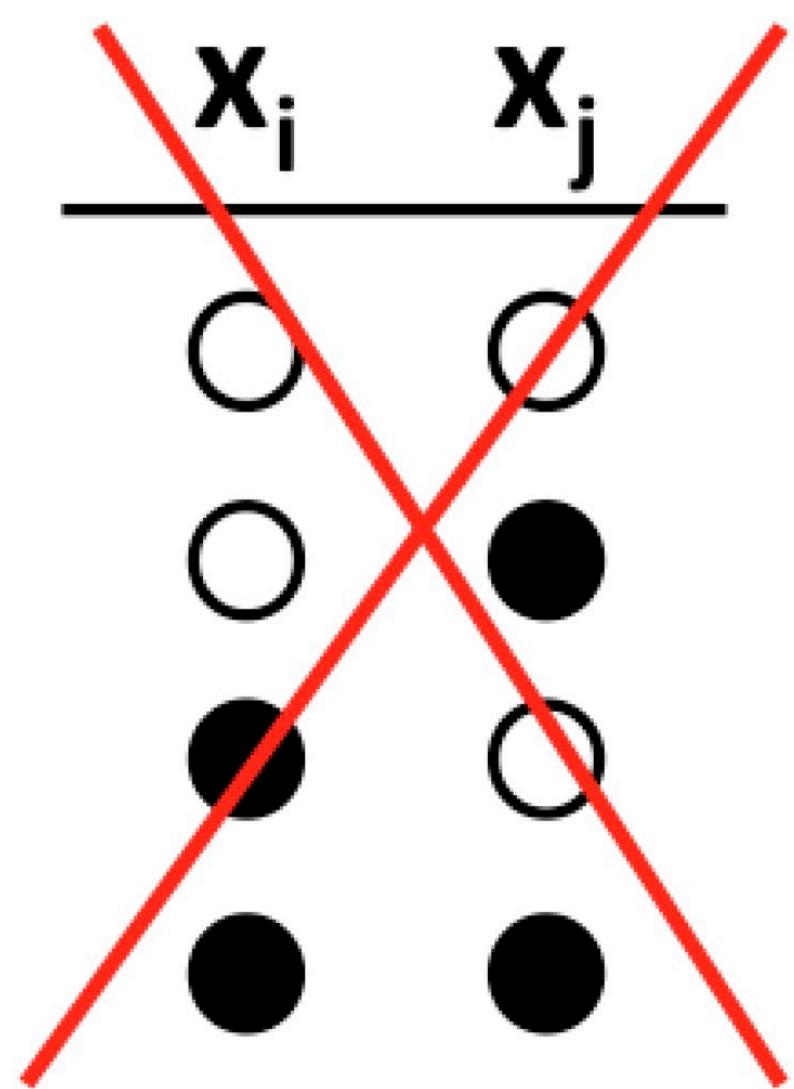
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
○	●	●

Puzzle

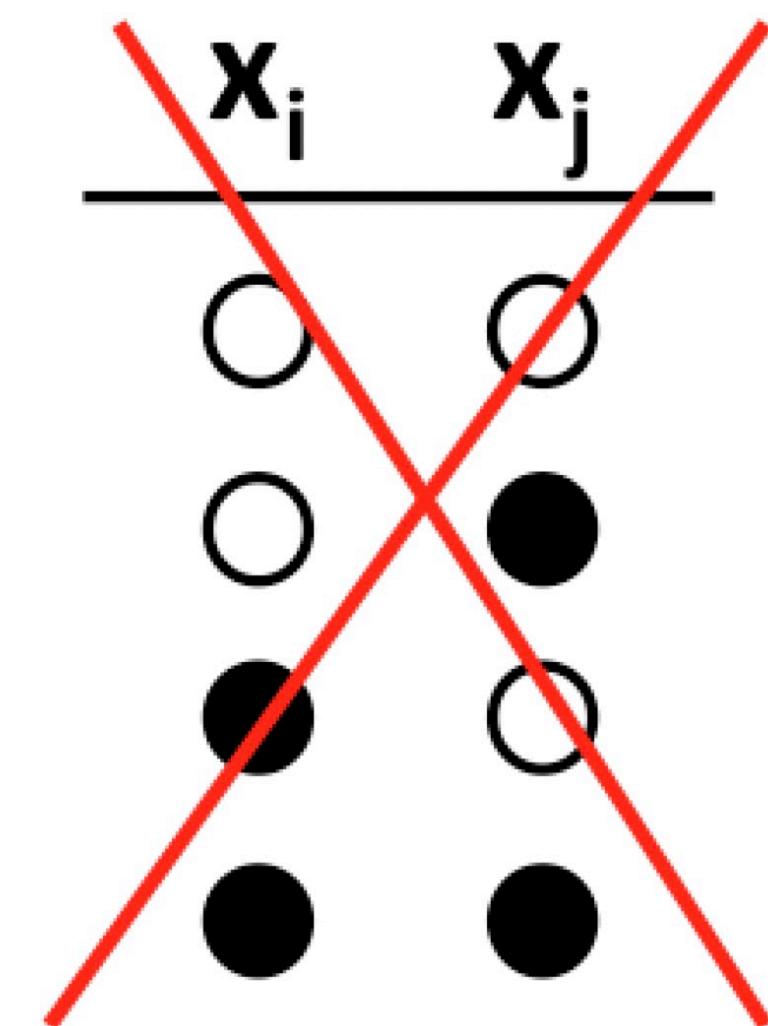
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
○	●	●

Puzzle

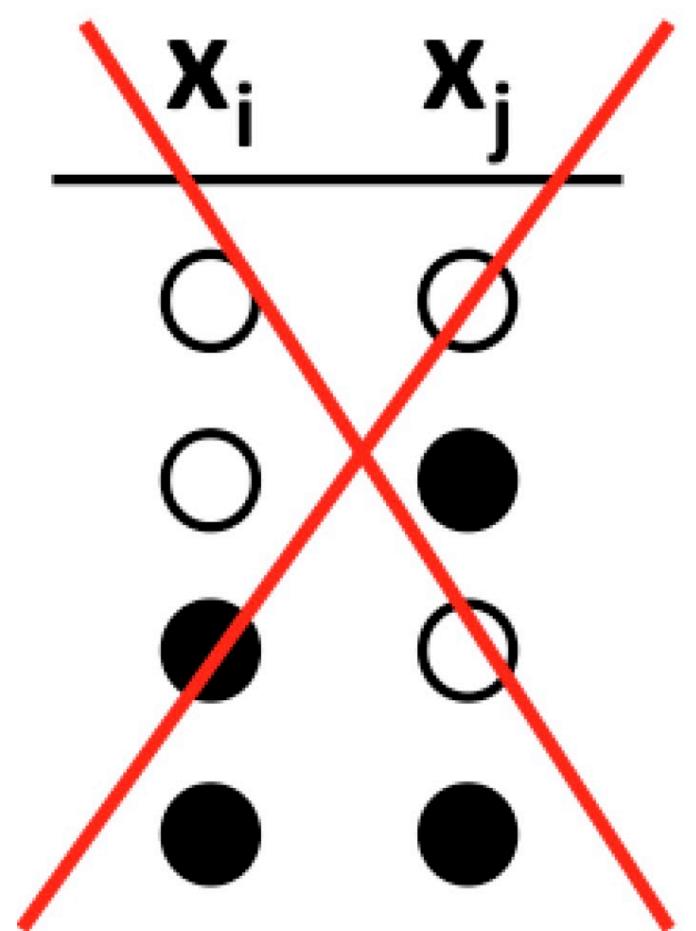
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○

Puzzle

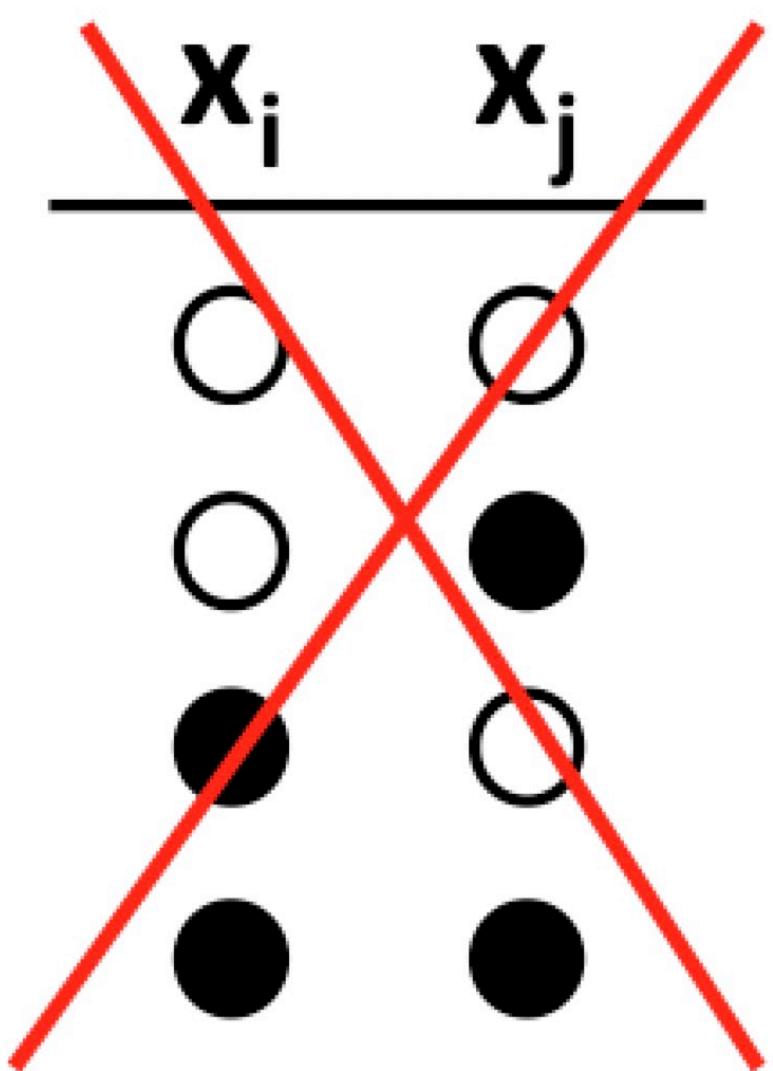
- Break point is $k = 2$



	x_1	x_2	x_3
1	○	○	○
2	○	○	●
3	○	●	○
4	●	○	○
5	●	○	●

Puzzle

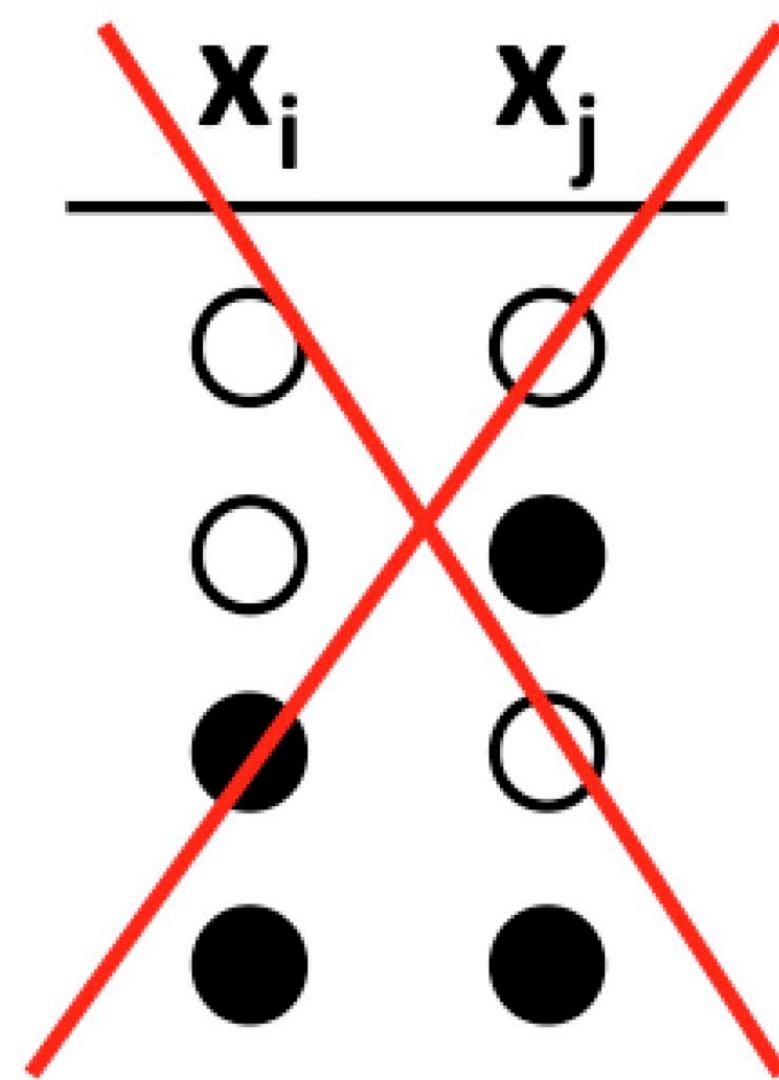
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○
●	○	●

Puzzle

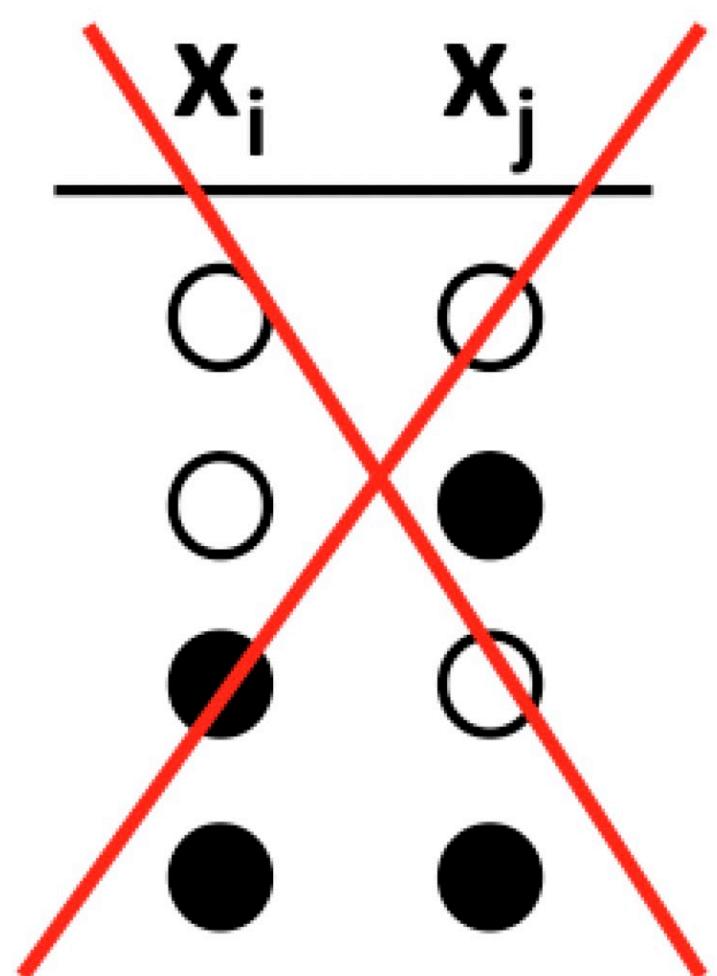
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○
●	●	○

Puzzle

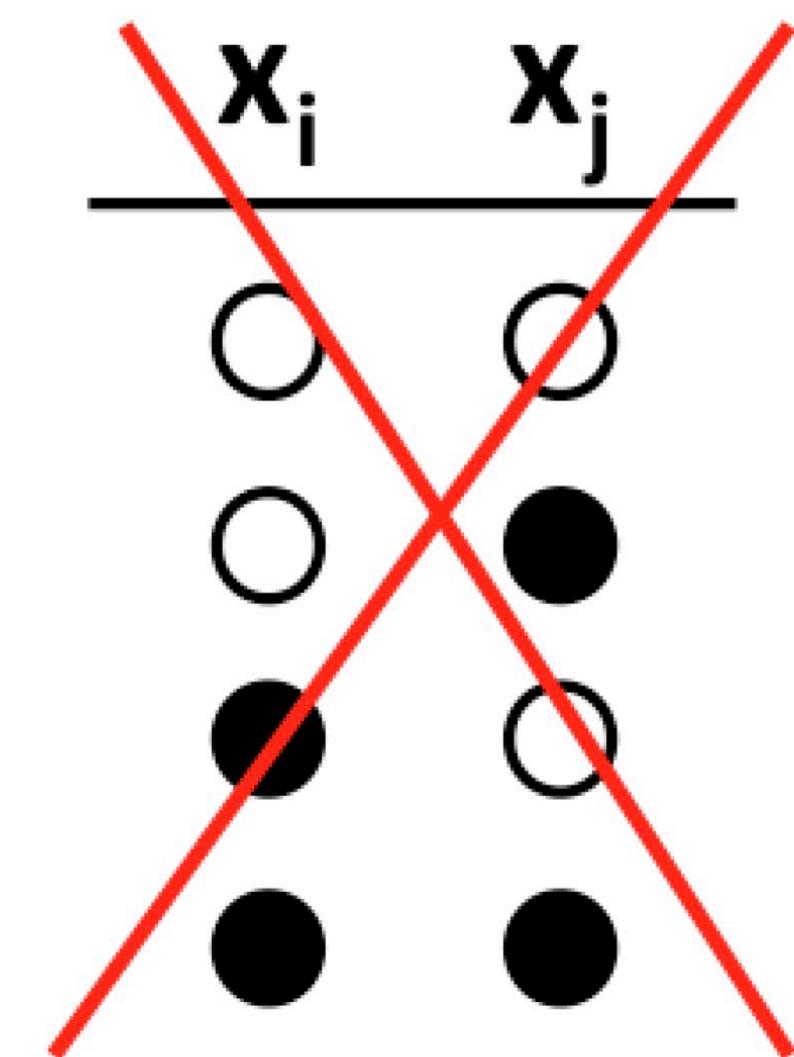
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○
●	●	○

Puzzle

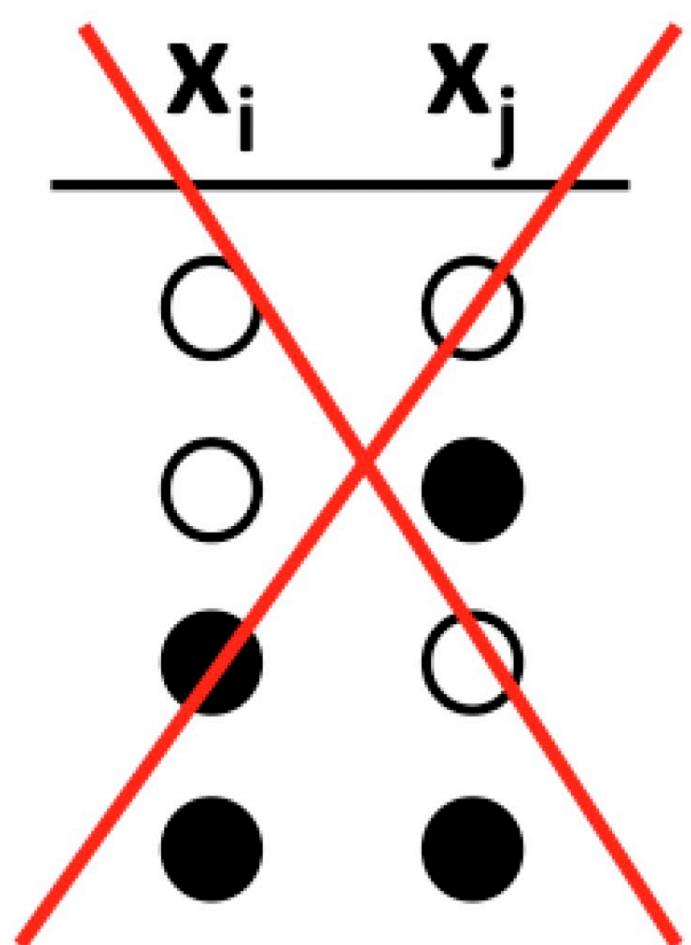
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○
●	●	●

Puzzle

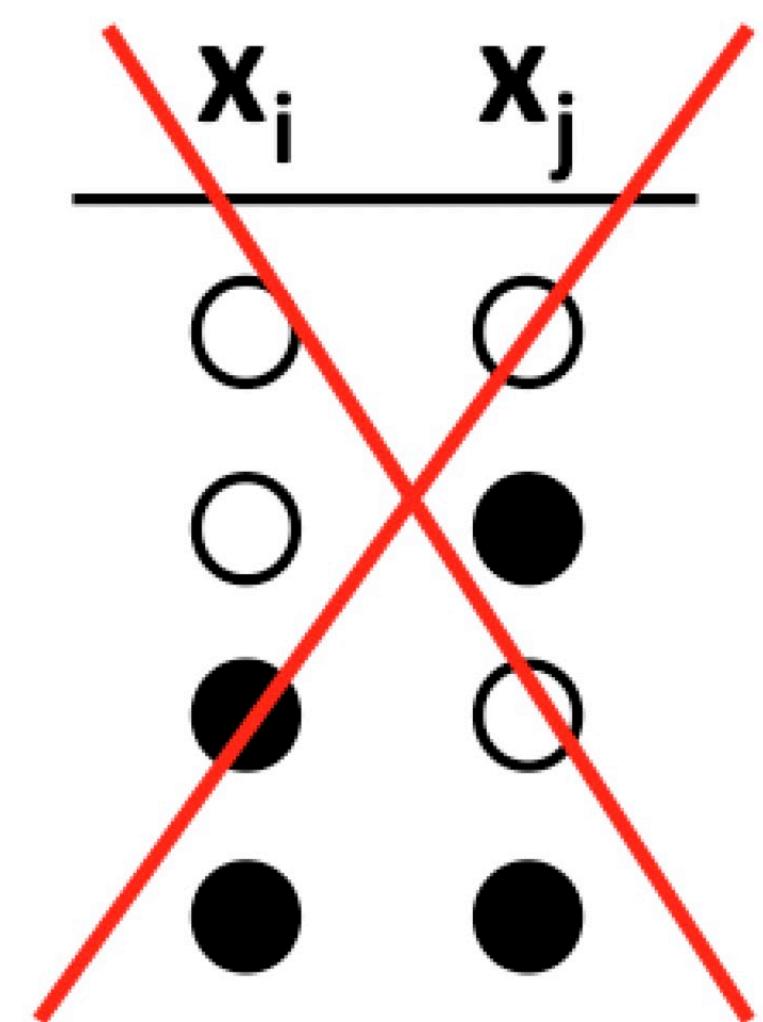
- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○
●	●	●

Puzzle

- Break point is $k = 2$



x_1	x_2	x_3
○	○	○
○	○	●
○	●	○
●	○	○

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

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

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

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

Recursive bound on $B(N, k)$

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

	# of rows	x_1	x_2	\dots	x_{N-1}	x_N
S_1	α	+1	+1	\dots	+1	+1
		-1	+1	\dots	+1	-1
		:	:	\vdots	:	:
		+1	-1	\dots	-1	-1
S_2	β	-1	+1	\dots	-1	+1
		+1	-1	\dots	+1	+1
		-1	-1	\dots	+1	+1
		:	:	\vdots	:	:
S_2	β	+1	-1	\dots	+1	-1
		-1	-1	\dots	+1	-1
		:	:	\vdots	:	:
		+1	-1	\dots	+1	-1
S_2^-	β	-1	-1	\dots	-1	-1
		+1	-1	\dots	+1	-1
		-1	-1	\dots	-1	-1
		:	:	\vdots	:	:

Recursive bound on $B(N, k)$

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

	# 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^+	β	+1	-1	\dots	+1	+1
		-1	-1	\dots	+1	+1
		\vdots	\vdots	\vdots	\vdots	\vdots
		+1	-1	\dots	+1	+1
S_2^-	β	-1	-1	\dots	-1	+1
		+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 $B(N, k)$

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

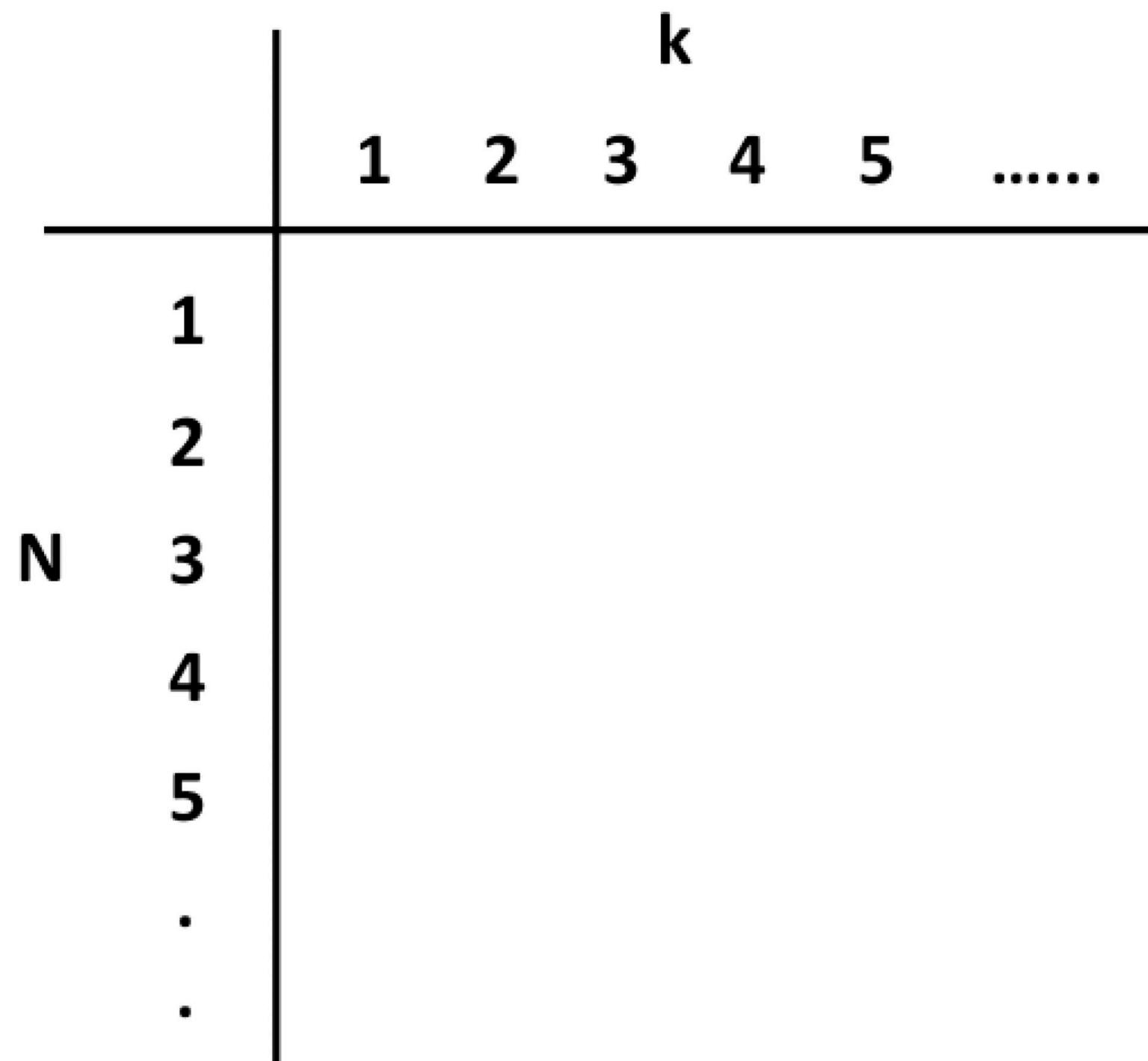
	# of rows	\mathbf{x}_1	\mathbf{x}_2	\dots	\mathbf{x}_{N-1}	\mathbf{x}_N
S_1	α	+1	+1	\dots	+1	+1
		-1	+1	\dots	+1	-1
		\vdots	\vdots	\vdots	\vdots	\vdots
		+1	-1	\dots	-1	-1
		-1	+1	\dots	-1	+1
S_2^+	β	+1	-1	\dots	+1	+1
		-1	-1	\dots	+1	+1
		\vdots	\vdots	\vdots	\vdots	\vdots
		+1	-1	\dots	+1	+1
		-1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

- $$B(N, k) = \alpha + \beta + \beta$$
- $\leq B(N - 1, k) + B(N - 1, k - 1)$
 - What's the upper bound for $B(N, k)$?

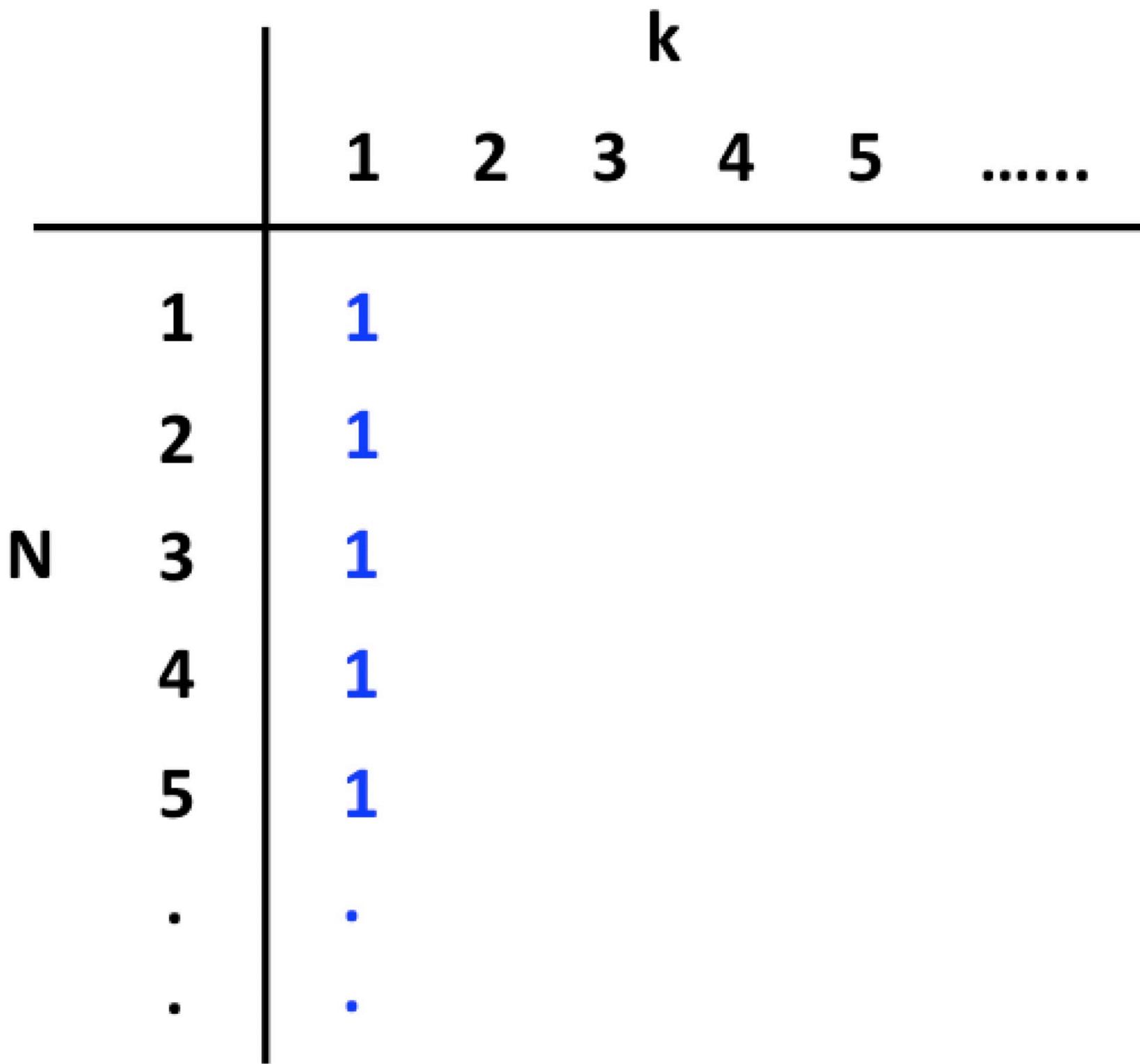
Recursive bound on $B(N, k)$

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



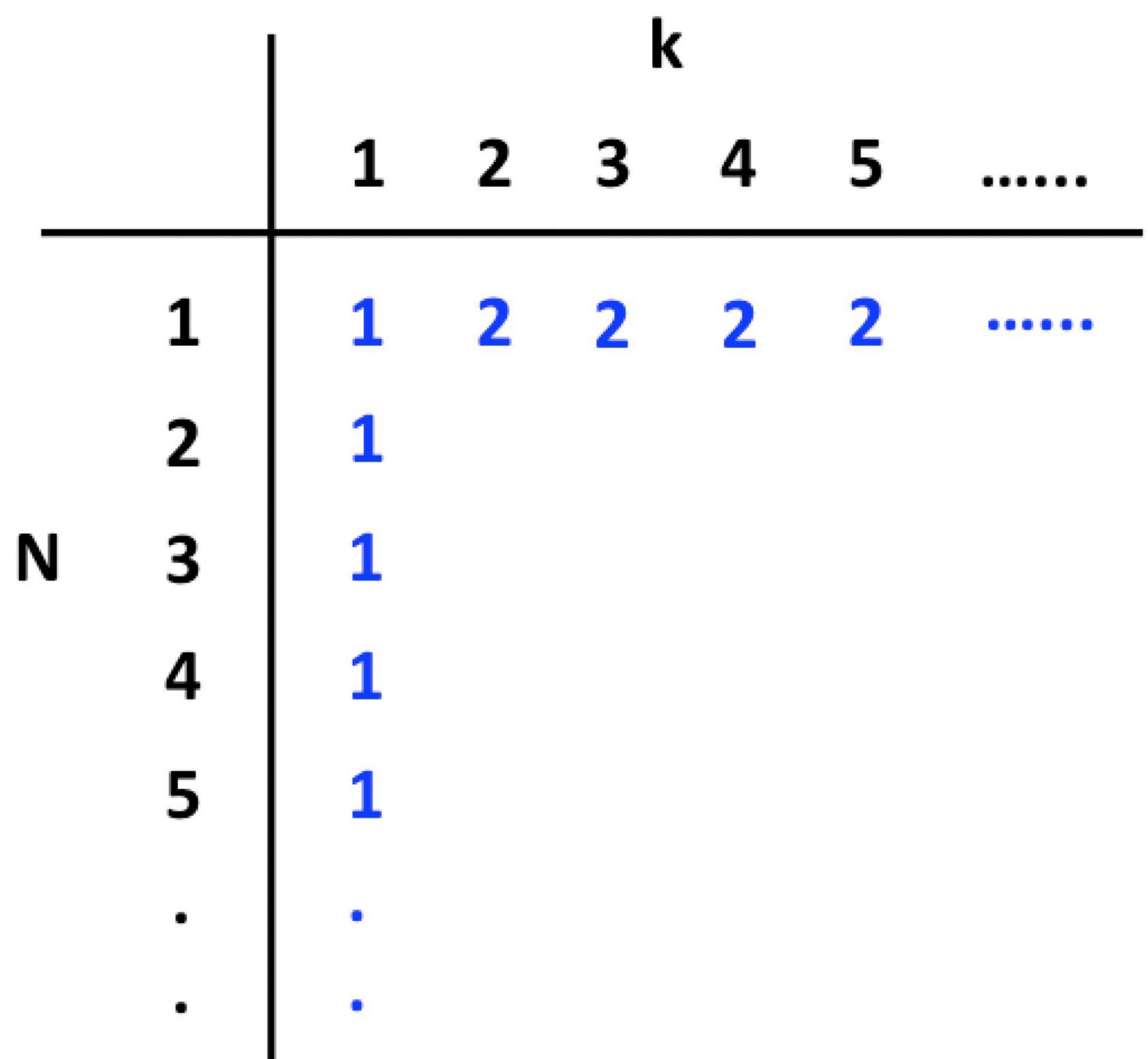
Recursive bound on $B(N, k)$

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



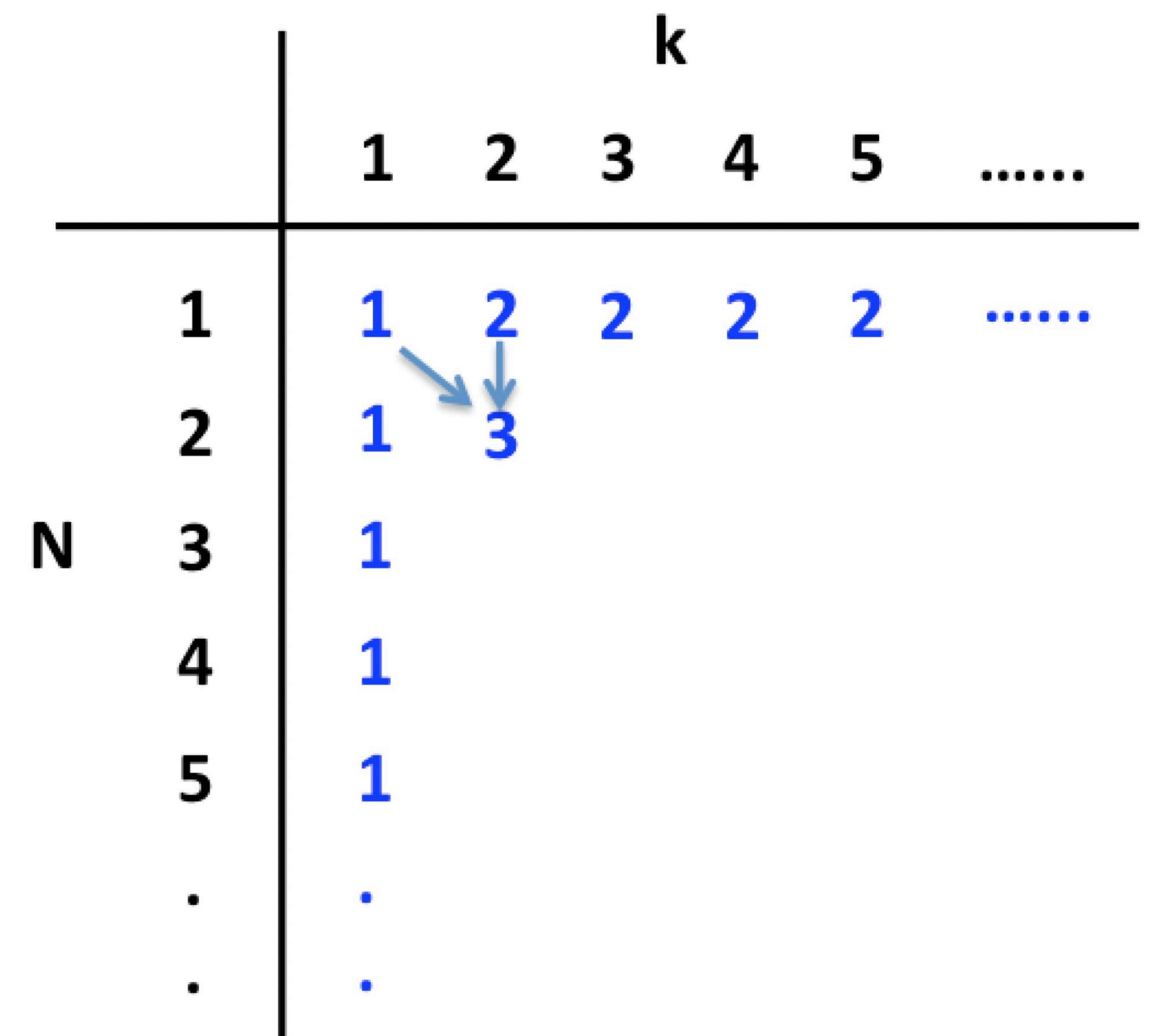
Recursive bound on $B(N, k)$

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



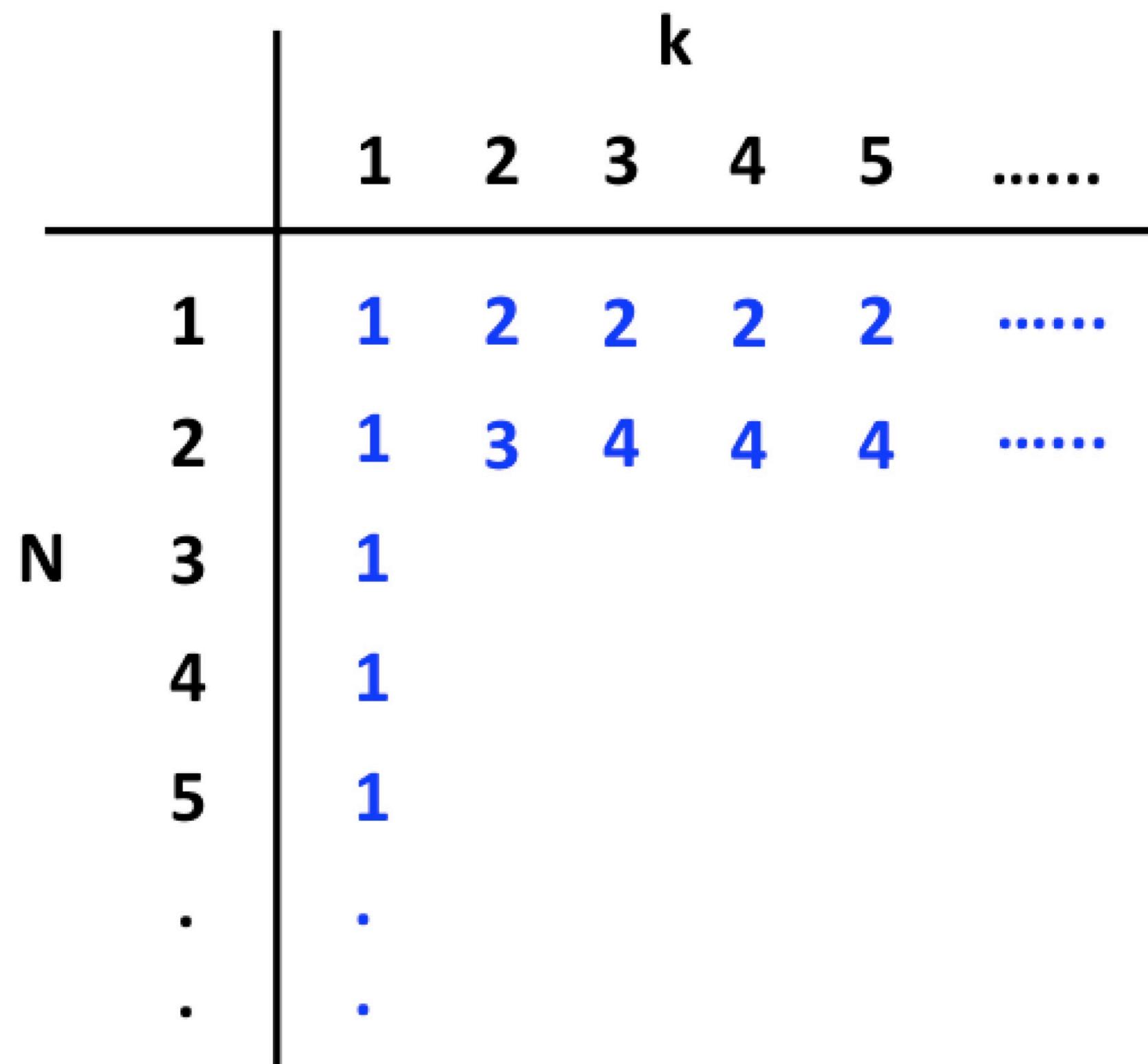
Recursive bound on $B(N, k)$

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



Recursive bound on $B(N, k)$

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



Recursive bound on $B(N, k)$

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

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

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

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

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

- $B(N, k)$ is upper bounded by $C(N, k)$
 - $C(N, 1) = 1, N = 1, 2, \dots$
 - $C(1, k) = 2, k = 2, 3, \dots$
 - $C(N, k) = C(N - 1, k) + C(N - 1, k - 1)$
- Sauer's Theorem: $C(N, k) = \sum_{i=0}^{k-1} \binom{N}{i}$
- Boundary conditions: (easy to check)
- Induction:
$$\sum_{i=0}^{k-1} \binom{N}{i} = \underbrace{\sum_{i=0}^{k-1} \binom{N-1}{i}}_{\text{select } < k \text{ from } N \text{ items}} + \underbrace{\sum_{i=0}^{k-2} \binom{N-1}{i}}_{\text{N-th item chosen}}$$

It is polynomial!

- For a given \mathcal{H} , the break point k is fixed:

$$m_{\mathcal{H}}(N) \leq \underbrace{\sum_{i=0}^{k-1} \binom{N}{i}}_{\text{Polynomial with degree } k-1}$$

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 positive rays: (break point $k = 2$)

- $m_{\mathcal{H}}(N) = N + 1$

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 $k = 4$)
 - $m_{\mathcal{H}}(N) = ?$

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 $k = 4$)

$$\bullet m_{\mathcal{H}}(N) \leq \frac{1}{6}N^3 + \frac{5}{6}N + 1$$

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

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

VC dimension

Definition

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

VC dimension

Definition

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

VC dimension

Definition

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

VC dimension

The growth function

- In terms of a break point k :

$$\bullet \quad m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

- In terms of the VC dimension d_{VC} :

$$\bullet \quad m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{VC}} \binom{N}{i}$$

VC dimension

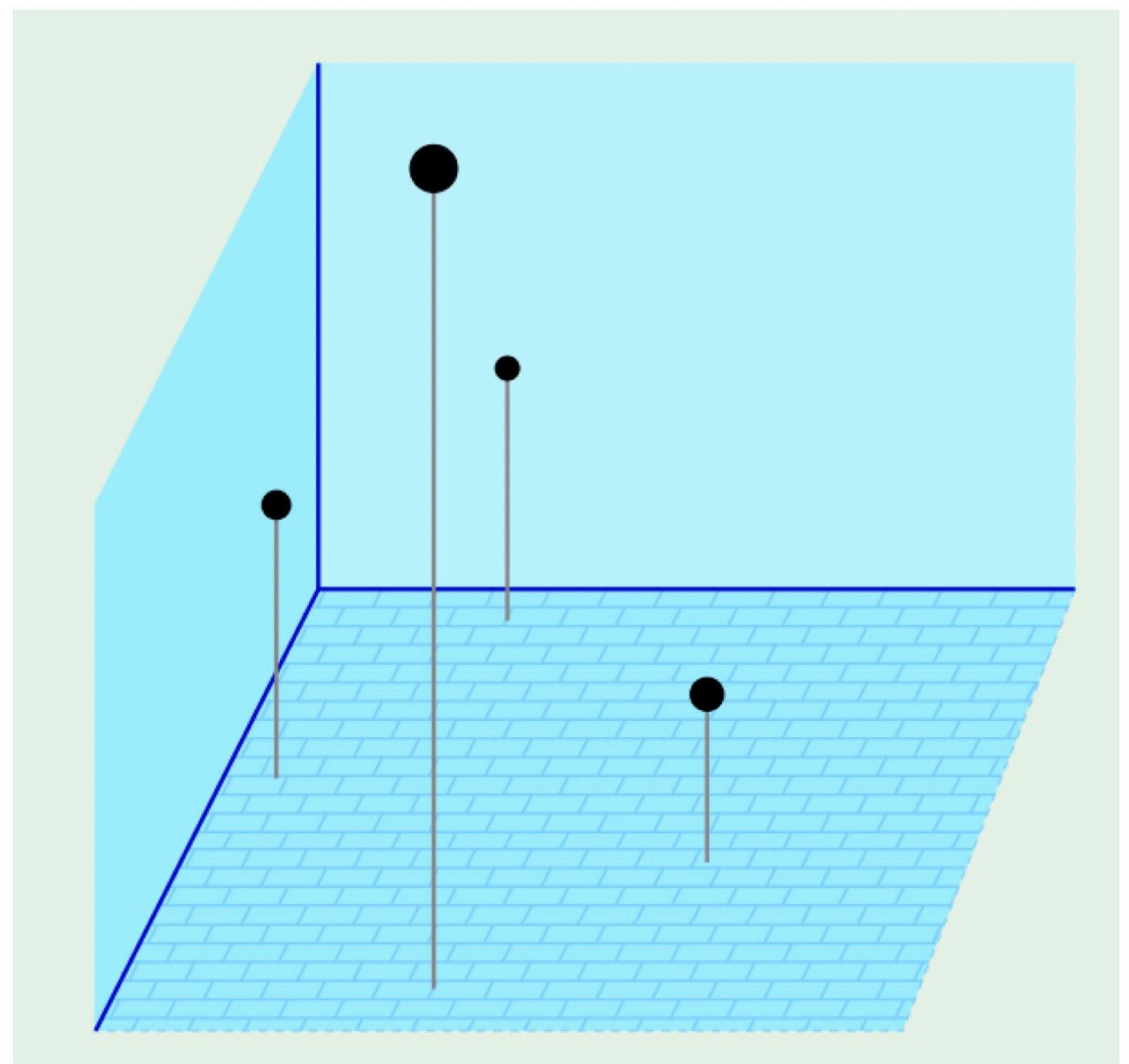
VC dimension of linear classifier

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

VC dimension

VC dimension of linear classifier

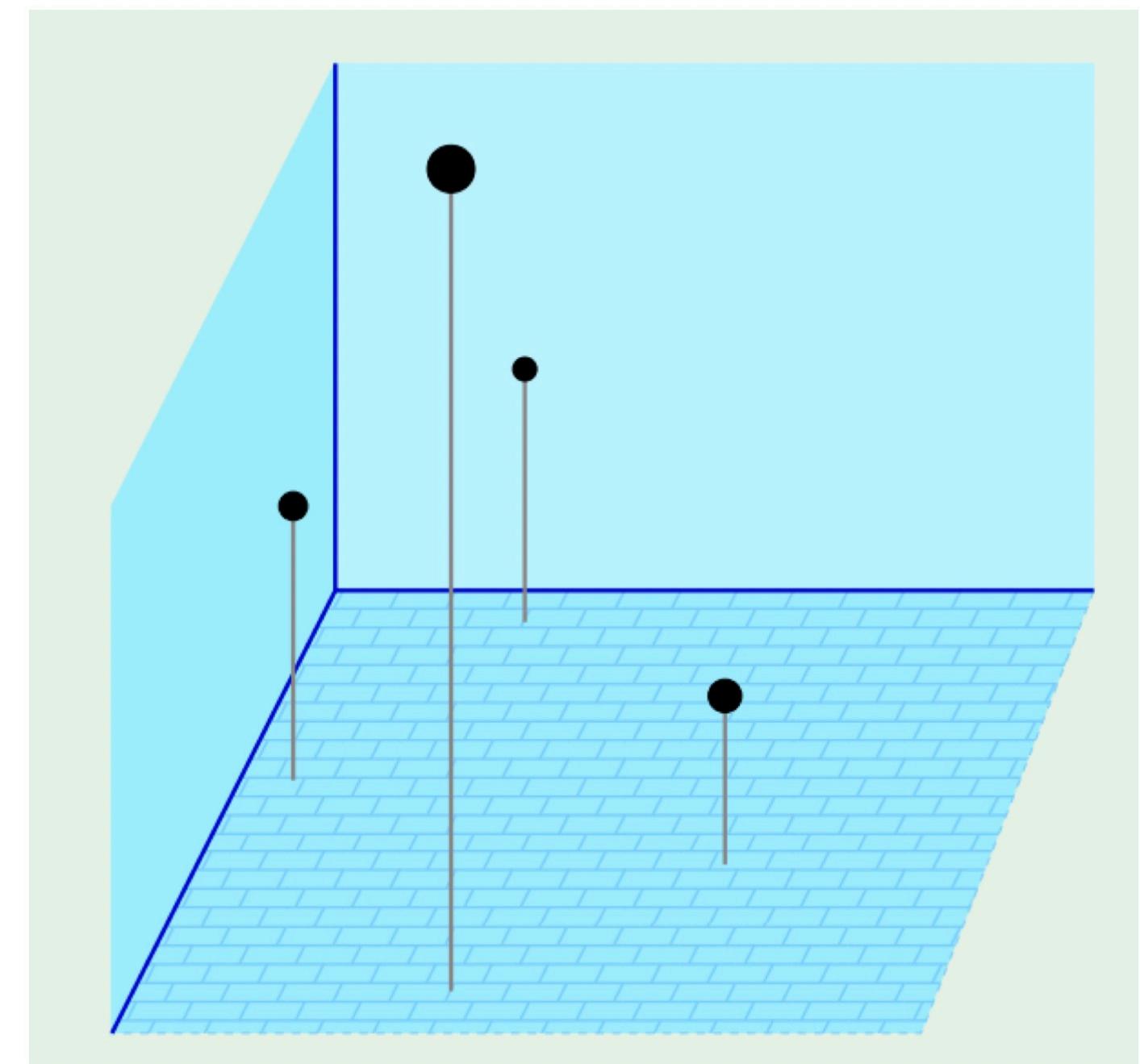
- For $d = 2$, $d_{VC} = 3$
- What if $d > 2$?



VC dimension

VC dimension of linear classifier

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



VC dimension

VC dimension of linear classifier

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

