Weak Convergence and Empirical Processes

Chapter 2.6: Uniform Entropy Numbers

Bart Eggen

Leiden University

February 3, 2020

What's on the menu today?

1. Introduction
2. VC-Classes
3. VC-Classes of Functions

Goal of this chapter

Uniform Entropy Number

$$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V, \quad 0 < \epsilon < 1$$

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Goal

Empirical process converges weakly to some limiting distribution for some class $\mathcal{F}$ of measurable functions

$\sup_Q N\left(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Goal

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \quad \text{in } \ell^\infty(\mathcal{F})$$

Recap

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \quad \text{in } \ell^\infty(\mathcal{F})$$

$\mathcal{F}$ is called a **P-Donsker class**.

Recap

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Theorem 2.5.2

Let $\mathcal{F}$ be a class of measurable functions that satisfies the uniform entropy bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty,$$

where the supremum is taken over all finitely discrete probability measures with $\|F\|_{Q,2} = \int F^2 dQ > 0$, where F is an envelope function for $\mathcal{F}$. Let the classes $\mathcal{F}_\delta = \left\{f - g : f, g \in \mathcal{F}, \|f - g\|_{P,2} < \delta\right\}$ and $\mathcal{F}_\infty^2$ be P-measurable for every $\delta > 0$. If $P^* F^2 < \infty$, then $\mathcal{F}$ is P-Donsker .

Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Covering number

The **covering number** $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$.

Uniform Entropy Bound

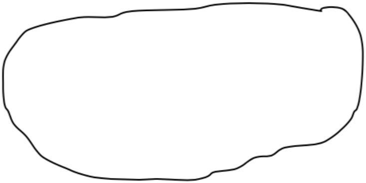
$$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \le K\left(\frac{1}{\epsilon}\right)^V$$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Covering number

The **covering number** $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$.



$\mathcal{F}$

Uniform Entropy Bound

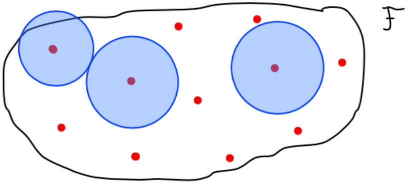
$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Covering number

The **covering number** $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$.



Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)}\,d\epsilon < \infty$$

Envelope function

An **envelope function** of a class $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$, for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$.

Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Envelope function

An **envelope function** of a class $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$, for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$.

Entropy number for $\epsilon > 1$

- For $\epsilon \geq 2$ we have $N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) = 1$

Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)\le K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)}\,d\epsilon<\infty$$

Envelope function

An **envelope function** of a class $\mathcal{F}$ is any function $x\mapsto F(x)$ such that $|f(x)|\le F(x)$, for every $x\in\mathcal{X}$ and $f\in\mathcal{F}$.

Entropy number for $\epsilon>1$

- For $\epsilon\ge 2$ we have $N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)=1$
- For $1\le\epsilon<2$ we have $N\left(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q)\right)\le 2$

Uniform Entropy Bound

$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^V$

Uniform Entropy Bound

$$\int_0^{1} \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Envelope function

An **envelope function** of a class $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$, for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$.

Entropy number for $\epsilon > 1$

- For $\epsilon \geq 2$ we have $N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) = 1$
- For $1 \leq \epsilon < 2$ we have $N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq 2$

Goal of this chapter

Uniform Entropy Bound

$$\int_0^1 \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Goal of this chapter

Uniform Entropy Bound

$$\int_0^1 \sup_Q \sqrt{\log N\left(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

Uniform entropy condition

$$\sup_Q \log N\left(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K \left(\frac{1}{\epsilon}\right)^{2-\delta}, \quad \delta > 0$$

Goal of this chapter

Uniform Entropy Bound

$$\int_0^1 \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}\, d\epsilon < \infty$$

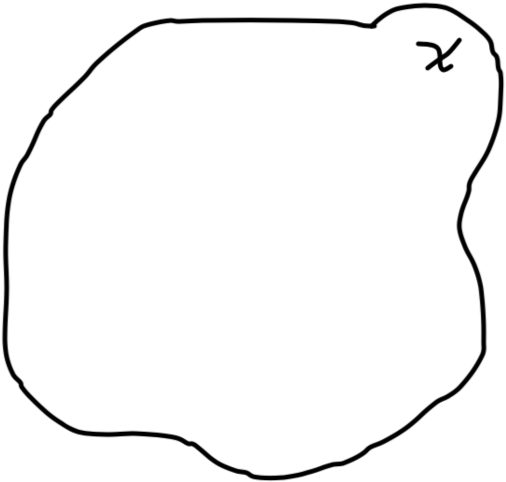
Uniform entropy condition

$$\sup_Q \log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^{2-\delta}, \quad \delta > 0$$

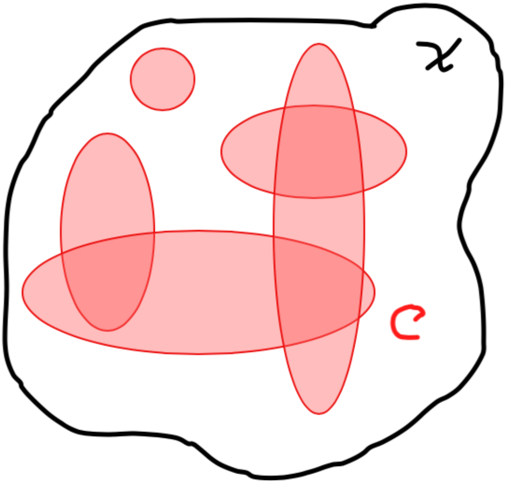
Our goal

$$\sup_Q N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^{V}, \quad \text{for some number } V$$

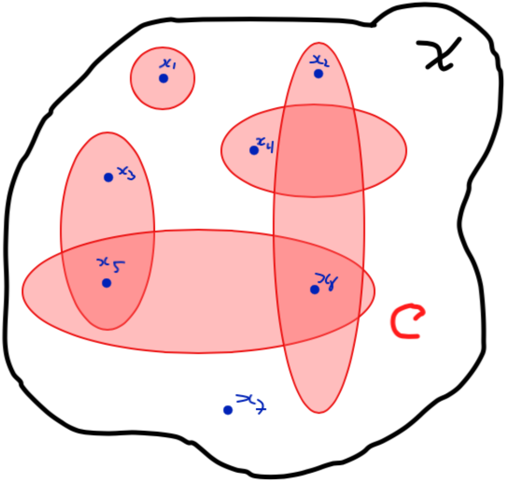
VC-Classes of Sets



VC-Classes of Sets

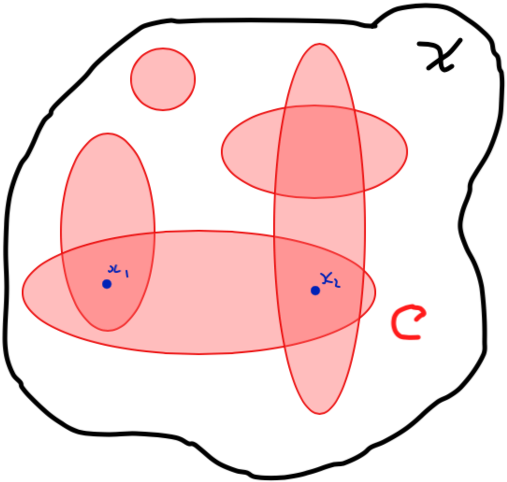


VC-Classes of Sets



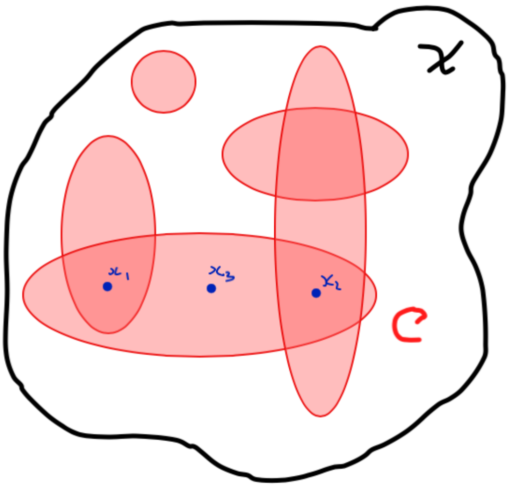
$\mathcal{C}$ **picks out** a subset $A \subseteq \{x_1, \ldots, x_n\}$ if there exists a $C \in \mathcal{C}$ with $C \cap \{x_1, \ldots, X_n\} = A$

VC-Classes of Sets



$\mathcal{C}$ **shatters** $\{x_1, \ldots, x_n\}$ if each of its 2^n subsets can be picked out

VC-Classes of Sets



The **VC-index** $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$

VC-Classes of Sets

VC-index

The **VC-index** of $\mathcal{C}$ is defined as

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\},$$

where

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\left\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\right\}.$$

A collection of measurable sets $\mathcal{C}$ is called a **VC-class** if its index is finite.

VC-Classes of Sets

VC-index

The **VC-index** of $\mathcal{C}$ is defined as

$$V(\mathcal{C}) = \inf\{n : \max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) < 2^n\},$$

where

$$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\left\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\right\}.$$

A collection of measurable sets $\mathcal{C}$ is called a **VC-class** if its index is finite.

Convention

The infimum over the empty set is taken to be infinity

Example

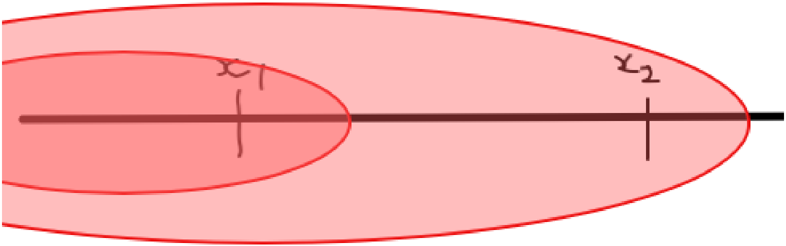
Real line

$$\mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}, \quad \{x_1, x_2\} \subset \mathbb{R}$$

Example

Real line

$$\mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}, \quad \{x_1, x_2\} \subset \mathbb{R}$$



Combinatorial results

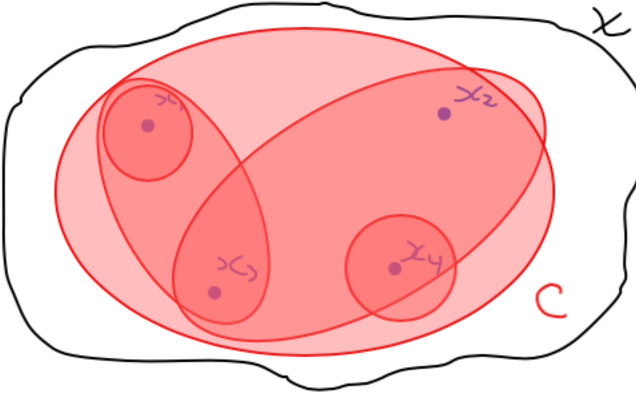
Lemma 2.6.2 (Sauer's lemma)

Let $\{x_1, \dots, x_n\} \subset \mathcal{X}$. Then the total number of subsets $\Delta_n(\mathcal{C}, x_1, \dots, x_n)$ picked out by $\mathcal{C}$ is bounded above by the number of subsets of $\{x_1, \dots, x_n\}$ shattered by $\mathcal{C}$.

Combinatorial results

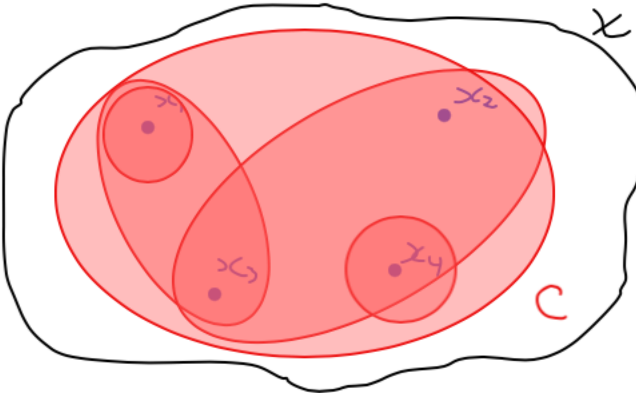
Lemma 2.6.2 (Sauer's lemma)

Let $\{x_1, \ldots, x_n\} \subset \mathcal{X}$. Then the total number of subsets $\Delta_n(\mathcal{C}, x_1, \ldots, x_n)$ picked out by $\mathcal{C}$ is bounded above by the number of subsets of $\{x_1, \ldots, x_n\}$ shattered by $\mathcal{C}$.



Proof Idea

$\Delta_n(\mathcal{C}, x_1, \ldots, x_n) = \#\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\}$



Without loss of generality:

$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$

Note that this way $|\mathcal{C}| = \Delta_n(\mathcal{C}, x_1, \ldots, x_n)$.

Proof Idea

Lemma 2.6.2 (Sauer's lemma)

Let $\{x_1, \ldots, x_n\} \subset \mathcal{X}$. Then $|\mathcal{C}|$ is bounded above by the number of subsets of $\{x_1, \ldots, x_n\}$ shattered by $\mathcal{C}$.

Definition

Call $\mathcal{C}$ **hereditary** if it has the property that $B \in \mathcal{C}$ whenever $B \subset C$, for a set $C \in \mathcal{C}$.

Note

Each of the sets in a hereditary set is shattered.

Proof Idea

Lemma 2.6.2 (Sauer's lemma)

Let $\{x_1, \ldots, x_n\} \subset \mathcal{X}$. Then $|\mathcal{C}|$ is bounded above by the number of subsets of $\{x_1, \ldots, x_n\}$ shattered by $\mathcal{C}$.

Definition

Call $\mathcal{C}$ **hereditary** if it has the property that $B \in \mathcal{C}$ whenever $B \subset C$, for a set $C \in \mathcal{C}$.

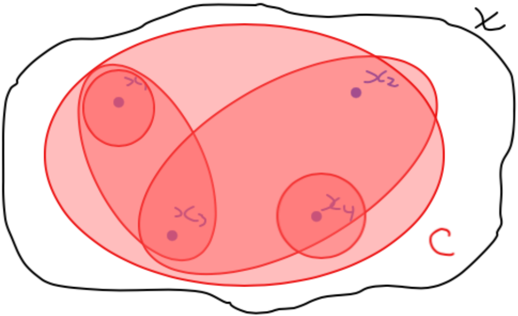
Note

Each of the sets in a hereditary set is shattered.

Idea

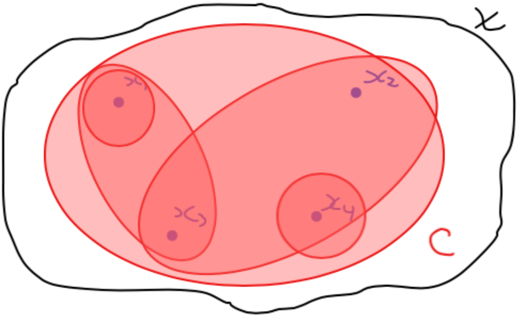
Transform each $\mathcal{C}$ to a hereditary collection, without changing the cardinality and without increasing the number of shattered sets.

Proof Idea



$$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

Proof Idea



$$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

Definition

For $1 \leq i \leq n$ and $C \in \mathcal{C}$

$$T_i(C) = \begin{cases} C - \{x_i\}, & C - \{x_i\} \notin \mathcal{C} \\ C, & \text{else.} \end{cases}$$

Proof Idea

Definition

For $1 \leq i \leq n$ and $C \in \mathcal{C}$

$$T_i(C) = \begin{cases} C - \{x_i\}, & C - \{x_i\} \notin \mathcal{C} \\ C, & \text{else.} \end{cases}$$

Note

- T_i is one-to-one, so $|\mathcal{C}| = |T_i(\mathcal{C})|$

Example

$$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

$$T_1(\mathcal{C}) = \{\{x_1\}, \{x_4\}, \{x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

Proof Idea

Definition

For $1 \leq i \leq n$ and $C \in \mathcal{C}$

$$T_i(C) = \begin{cases} C - \{x_i\}, & C - \{x_i\} \notin \mathcal{C} \\ C, & \text{else.} \end{cases}$$

Note

- T_i is one-to-one, so $|\mathcal{C}| = |T_i(\mathcal{C})|$
- Every subset $A \subset \{x_1, \ldots, x_n\}$ that is shattered by $T_i(\mathcal{C})$ is shattered by $\mathcal{C}$

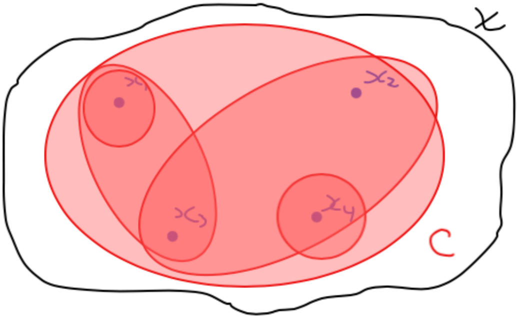
Example

$$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

$$T_1(\mathcal{C}) = \{\{x_1\}, \{x_4\}, \{x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

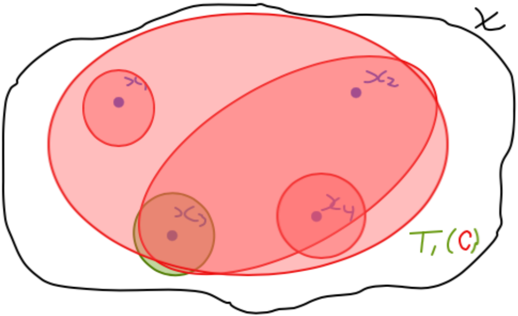
$T_1(\mathcal{C})$ shatters $\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}, \{x_3.x_4\}$

Proof Idea



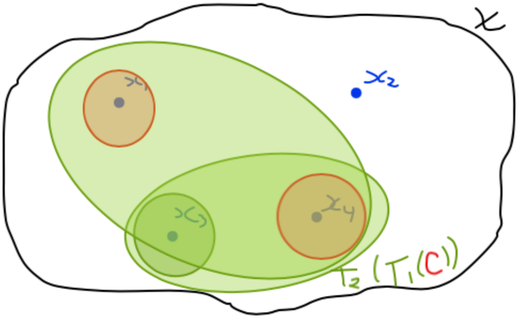
$$\mathcal{C} = \{\{x_1\}, \{x_4\}, \{x_1, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

Proof Idea



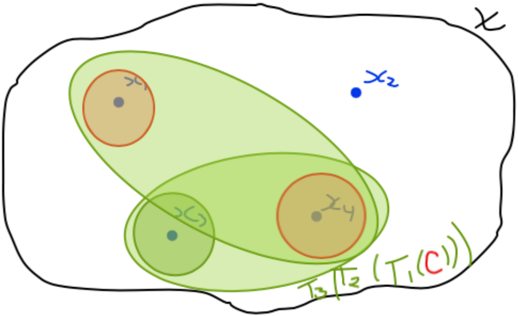
$$T_1(\mathcal{C}) = \{\{x_1\}, \{x_4\}, \{x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_2, x_3, x_4\}\}$$

Proof Idea



$$(T_2 \circ T_1)(\mathcal{C}) = \{\{x_1\}, \{x_4\}, \{x_3\}, \{x_3, x_4\}, \{x_1, x_3, x_4\}\}$$

Proof Idea



$$(T_3 \circ T_2 \circ T_1)(\mathcal{C}) = \{\{x_1\}, \{x_4\}, \{x_3\}, \{x_3, x_4\}, \{x_1, x_4\}\}$$

Corollary

Corollary 2.6.3

For a VC-class of sets of index $V(\mathcal{C})$, one has

$$\max_{x_1,\dots,x_n} \Delta_n(\mathcal{C}, x_1, \dots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j}.$$

Consequently, the numbers on the left side grow polynomially of order at most $O\left(n^{V(\mathcal{C})-1}\right)$ as $n \to \infty$.

Most important theorem

Theorem 2.6.4

There exists a universal constant K such that for any VC-class $\mathcal{C}$ of sets, any probability measure Q, any $r \geq 1$, and $0 < \epsilon < 1$,

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}$$

Proof idea

1. Prove it for $r = 1$ and use a simple argument to show for $r > 1$
2. Use problem 2.6.3 to show it is enough to check if it holds for empirical type measures
3. Use some difficult arguments on the n-dimensional hypercube to obtain a bound on the packing number

Proof (idea)

Proof parts

1. Simple argument for the covering number for general $r > 1$ if the inequality holds for $r = 1$

Proof (idea)

Proof parts

1. Simple argument for the covering number for general $r > 1$ if the inequality holds for $r = 1$

Suppose

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq f(\epsilon).$$

Proof (idea)

Proof parts

1. Simple argument for the covering number for general $r > 1$ if the inequality holds for $r = 1$

Suppose

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq f(\epsilon).$$

Note that for any $C, D \in \mathcal{C}$ we have

$$\begin{aligned}
\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,r} &= \left(\int |\mathbb{1}_C - \mathbb{1}_D|^r dQ\right)^{1/r} \\
&= \left(\int |\mathbb{1}_C - \mathbb{1}_D| dQ\right)^{1/r} \\
&= \left(\int \mathbb{1}_{(C\cup D)\setminus(C\cap D)} dQ\right)^{1/r} \\
&= Q^{1/r}(C \triangle D)
\end{aligned}$$

Proof (idea)

Proof parts

1. Simple argument for the covering number for general $r > 1$ if the inequality holds for $r = 1$

Suppose

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq f(\epsilon).$$

Then

$$N(\epsilon, \mathcal{C}, L_r(Q)) = N(\epsilon^r, \mathcal{C}, L_1(Q)) \leq f(\epsilon^r)$$

Prrof (idea)

Proof parts

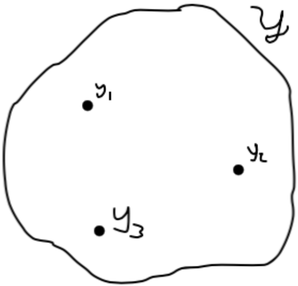
2. Use problem 2.6.3 to show that it is enough to check if it holds for empirical type measures

Prrof (idea)

Suppose Q is a measure on a finite set of points $y_1, \ldots, y_k$ such that $Q(\{y_i\}) = l_i/n$ for integers $l_1, \ldots, l_k$ that add up to n. We assume again that each set in $\mathcal{C}$ is a subset of these points.

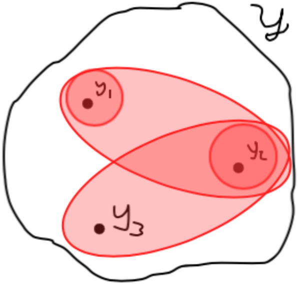
Prrof (idea)

$$Q(\{y_1\}) = \frac{1}{5} \qquad Q(\{y_2\}) = \frac{2}{5} \qquad Q(\{y_3\}) = \frac{2}{5}$$



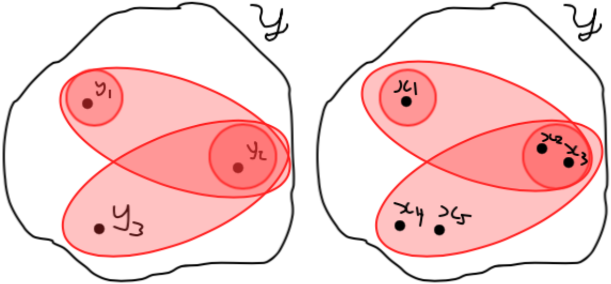
Prrof (idea)

$$Q(\{y_1\}) = \frac{1}{5} \qquad Q(\{y_2\}) = \frac{2}{5} \qquad Q(\{y_3\}) = \frac{2}{5}$$



$$\mathcal{C} = \{\{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_2, y_3\}\}$$

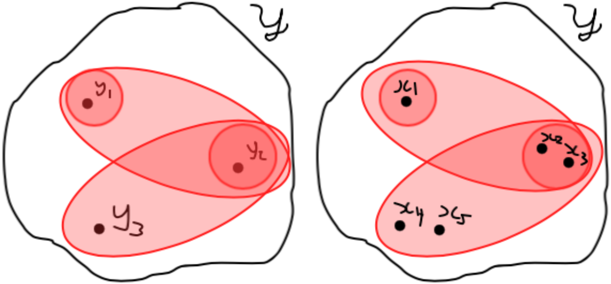
Prrof (idea)



$$\mathcal{C} = \{\{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_2, y_3\}\}$$

$$\widetilde{\mathcal{C}} = \{\{x_1\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4, x_5\}\}$$

Prrof (idea)



$$\mathcal{C} = \{\{y_1\}, \{y_2\}, \{y_1, y_2\}, \{y_2, y_3\}\}$$

$$\widetilde{\mathcal{C}} = \{\{x_1\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4, x_5\}\}$$

Note

The VC-index of $\widetilde{\mathcal{C}}$ is the same as $\mathcal{C}$

Prrof (idea)

Covering number for the new collection $\widetilde{\mathcal{C}}$

Let $C, D \in \mathcal{C}$, then we have $\widetilde{C}, \widetilde{D} \in \widetilde{\mathcal{C}}$. Take $\widetilde{Q}(\{x_i\}) = \frac{1}{n}$ for all $1 \leq i \leq n$. Then

$$Q(C \triangle D) = \widetilde{Q}(\widetilde{C} \triangle \widetilde{D})$$

Prrof (idea)

Covering number for the new collection $\widetilde{\mathcal{C}}$

Let $C, D \in \mathcal{C}$, then we have $\widetilde{C}, \widetilde{D} \in \widetilde{\mathcal{C}}$. Take $\widetilde{Q}(\{x_i\}) = \frac{1}{n}$ for all $1 \leq i \leq n$. Then

$$Q(C \triangle D) = \widetilde{Q}(\widetilde{C} \triangle \widetilde{D})$$

Remember

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} = Q(C \triangle D)$$

Prrof (idea)

Covering number for the new collection $\widetilde{\mathcal{C}}$

Let $C, D \in \mathcal{C}$, then we have $\widetilde{C}, \widetilde{D} \in \widetilde{\mathcal{C}}$. Take $\widetilde{Q}\left(\{x_i\}\right) = \frac{1}{n}$ for all $1 \leq i \leq n$. Then

$$Q(C \triangle D) = \widetilde{Q}(\widetilde{C} \triangle \widetilde{D})$$

Remember

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} = Q(C \triangle D)$$

Conclusion

$$N\left(\epsilon, \widetilde{\mathcal{C}}, L_1(\widetilde{Q})\right) = N\left(\epsilon, \mathcal{C}, L_1(Q)\right)$$

Proof (idea)

Proof parts

3. Use arguments on the n-dimensional hypercube to obtain a bound on the packing number

Proof (idea)

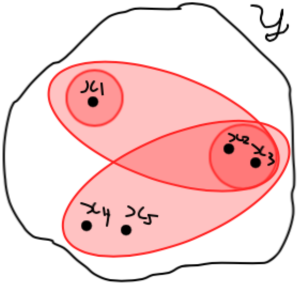
Proof parts

3. Use arguments on the n-dimensional hypercube to obtain a bound on the packing number

Note

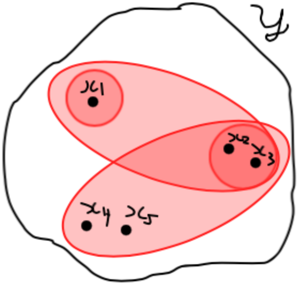
$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q))$$

Proof (idea)



$$\mathcal{C} = \{\{x_1\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4, x_5\}, \emptyset\}$$

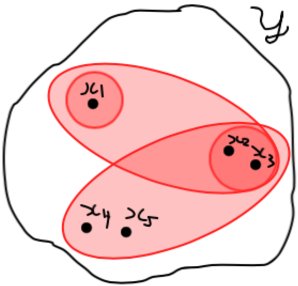
Proof (idea)



$$\mathcal{C} = \{\{x_1\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4, x_5\}, \emptyset\}$$

$$C_1 = \begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix} \quad C_2 = \begin{pmatrix}0\\1\\1\\0\\0\end{pmatrix} \quad C_3 = \begin{pmatrix}1\\1\\1\\0\\0\end{pmatrix} \quad C_4 = \begin{pmatrix}0\\1\\1\\1\\1\end{pmatrix} \quad C_5 = \begin{pmatrix}0\\0\\0\\0\\0\end{pmatrix}$$

Proof (idea)



$$\mathcal{C} = \{\{x_1\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4, x_5\}, \emptyset\}$$

Note

The collection $\mathcal{C}$ can be identified with a subset $\mathcal{Z}$ of the vertices of the n-dimensional hypercube $[0, 1]^n$

Proof (idea)

$$\mathcal{C} \longrightarrow \mathcal{Z} \longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Proof (idea)

$$\mathcal{C} \longrightarrow \mathcal{Z} \longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$I = \{1, 2\} \qquad \mathcal{Z}_I := \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Proof (idea)

$$\mathcal{C} \longrightarrow \mathcal{Z} \longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$I = \{1, 2\} \qquad\qquad \mathcal{Z}_I := \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Note

- $\mathcal{Z}_I$ corresponds to the collection of $C \cap \{x_i : i \in I\}$
- $\{x_1 : i \in I\}$ is shattered of $\mathcal{Z}_I$ consist of all $2^{|I|}$ possible columns
- This is only possible for $|I| < V(\mathcal{C})$, so define $V = V(\mathcal{C}) - 1$

Proof (idea)

Note

We transformed our problem from $\mathcal{C}$ to $\mathcal{Z}$. However, we also need to transform our metric!

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} = Q(C \triangle D)$$

Proof (idea)

Note

We transformed our problem from $\mathcal{C}$ to $\mathcal{Z}$. However, we also need to transform our metric!

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} = Q(C \triangle D)$$

Hamming metric

Let the vertices $w, z \in \mathcal{Z}$ correspond to the sets $C, D \in \mathcal{C}$ respectively. The *Hamming metric* on $\mathcal{Z}$ is defined by

$$d(w,z) = \frac{1}{n}\sum_{j=1}^{n} |w_j - z_j| = \frac{1}{n}\sum_{j=1}^{n} (w_j - z_j)^2, \quad z, w \in \mathcal{Z}.$$

With this metric we have $Q(C \triangle D) = d(w,z)$.

Proof (idea)

Note

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q))$$

Proof (idea)

Note

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q))$$

ϵ-seperation

Fix a maximal ϵ-seperated collection of sets $C \in \mathcal{C}$. For simplicity of notation we assume that $\mathcal{C}$ is ϵ-seperated, so for $C, D \in \mathcal{C}$

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} > \epsilon$$

Proof (idea)

Note

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q))$$

ϵ-seperation

Fix a maximal ϵ-seperated collection of sets $C \in \mathcal{C}$. For simplicity of notation we assume that $\mathcal{C}$ is ϵ-seperated, so for $C, D \in \mathcal{C}$

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} > \epsilon$$

Note

In this way, $\mathcal{Z}$ is also ϵ-seperated using the Hamming distance. For $w, z \in \mathcal{Z}$ corresponding to $C, D \in \mathcal{C}$ we have

$$d(w, z) > \epsilon.$$

Proof (idea)

Note

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq D(\epsilon, \mathcal{C}, L_1(Q)) \leq |\mathcal{Z}|$$

ϵ-seperation

Fix a maximal ϵ-seperated collection of sets $C \in \mathcal{C}$. For simplicity of notation we assume that $\mathcal{C}$ is ϵ-seperated, so for $C, D \in \mathcal{C}$

$$\|\mathbb{1}_C - \mathbb{1}_D\|_{Q,1} > \epsilon$$

Note

In this way, $\mathcal{Z}$ is also ϵ-seperated using the Hamming distance. For $w, z \in \mathcal{Z}$ corresponding to $C, D \in \mathcal{C}$ we have

$$d(w, z) > \epsilon.$$

Proof

Introducing random variables

Let Z be a random variable with a discrete uniform distribution on the set $\mathcal{Z}$ in $[0,1]^n$, so

$$P(Z = z) = \frac{1}{|\mathcal{Z}|}, \quad z \in \mathcal{Z}.$$

Proof

Introducing random variables

Let Z be a random variable with a discrete uniform distribution on the set $\mathcal{Z}$ in $[0,1]^n$, so

$$P(Z = z) = \frac{1}{|\mathcal{Z}|}, \quad z \in \mathcal{Z}.$$

Fix integer

Fix integer $V \leq m < n$. Let $I \subseteq \{1, 2, \ldots, n\}$ such that $|I| = m + 1$.

Proof

Introducing random variables

Let Z be a random variable with a discrete uniform distribution on the set $\mathcal{Z}$ in $[0,1]^n$, so

$$P(Z = z) = \frac{1}{|\mathcal{Z}|}, \quad z \in \mathcal{Z}.$$

Fix integer

Fix integer $V \leq m < n$. Let $I \subseteq \{1, 2, \ldots, n\}$ such that $|I| = m + 1$.

Lemma 2.6.6

Let Z be an arbitrary random vector taking values in $\mathcal{Z} \subset \{0,1\}^n$ that corresponds to a VC-class $\mathcal{C}$ of subsets of a set of points $\{x_1, \ldots, x_n\}$. Then

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$$

Proof

Introducing random variables

Let Z be a random variable with a discrete uniform distribution on the set $\mathcal{Z}$ in $[0,1]^n$, so

$$P(Z = z) = \frac{1}{|\mathcal{Z}|}, \quad z \in \mathcal{Z}.$$

Fix integer

Fix integer $V \leq m < n$. Let $I \subseteq \{1, 2, \ldots, n\}$ such that $|I| = m + 1$.

Our case

Apply the lemma to Z_I:

$$\sum_{i \in I} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_{I - \{i\}}\right)\right] \leq V$$

Proof

$$\sum_{i \in I} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_{I-\{i\}}\right)\right] \leq V$$

Proof

$$\sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I| = m+1}} \sum_{i \in I} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_{I - \{i\}}\right)\right] \leq \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I| = m+1}} V$$

Proof

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \sum_{i \notin J} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_J\right)\right] \leq \binom{n}{m+1} V$$

Proof

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J\right)\right] \leq \binom{n}{m+1} V$$

Proof

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J\right)\right] \leq \binom{n}{m+1} V$$

Proof

Remember

Z is uniformly distributed on the set $\mathcal{Z}$

Proof

Remember

Z is uniformly distributed on the set $\mathcal{Z}$

Consequence

$Z|Z_J = s$ is uniformly distributed over the set of columns $z \in \mathcal{Z}$ with $z_J = s$. Call N_s the number of these columns.

Proof

Remember

Z is uniformly distributed on the set $\mathcal{Z}$

Consequence

$Z|Z_J = s$ is uniformly distributed over the set of columns $z \in \mathcal{Z}$ with $z_J = s$. Call N_s the number of these columns.

Define

Let W and $\tilde{W}$ be independent random vectors defined on a common probability space distributed uniformly over these columns.

Proof

Remember

$\mathcal{Z}$ is ϵ-seperated

Proof

Remember

$\mathcal{Z}$ is ϵ-seperated

Consequence

If $W \neq \tilde{W}$, we have $d(W, \tilde{W}) > \epsilon$. This event happens with probability $1 - 1/N_s$. Else, $d(W, \tilde{W}) = 0$.

Proof

Distance

$$d(W, \tilde{W}) > \epsilon \text{ with probability } 1 - 1/N_s$$

Proof

Distance

$$d(W, \tilde{W}) > \epsilon \text{ with probability } 1 - 1/N_s$$

$$\begin{aligned}
\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J = s\right) &= \frac{1}{2} \sum_{i=1}^{n} \left(2\mathbb{E}\left[W_i^2\right] - 2\mathbb{E}\left[W_i\right]^2\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \left(\mathbb{E}\left[W_i^2\right] + \mathbb{E}\left[\tilde{W}_i^2\right] - 2\mathbb{E}\left[W_i\right]\mathbb{E}\left[\tilde{W}_i\right]\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[\left(W_i - \tilde{W}_i\right)^2\right]
\end{aligned}$$

Proof

Distance

$$d(W, \tilde{W}) > \epsilon \text{ with probability } 1 - 1/N_s$$

$$\begin{aligned}
\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J = s\right) &= \frac{1}{2} \sum_{i=1}^{n} \left(2\mathbb{E}\left[W_i^2\right] - 2\mathbb{E}\left[W_i\right]^2\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \left(\mathbb{E}\left[W_i^2\right] + \mathbb{E}\left[\tilde{W}_i^2\right] - 2\mathbb{E}\left[W_i\right] \mathbb{E}\left[\tilde{W}_i\right]\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[\left(W_i - \tilde{W}_i\right)^2\right]
\end{aligned}$$

Remember

$$d(w, z) = \frac{1}{n} \sum_{j=1}^{n} |w_j - z_j| = \frac{1}{n} \sum_{j=1}^{n} (w_j - z_j)^2, \quad z, w \in \mathcal{Z}.$$

Proof

Distance

$$d(W, \tilde{W}) > \epsilon \text{ with probability } 1 - 1/N_s$$

$$\begin{aligned}
\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J = s\right) &= \frac{1}{2} \sum_{i=1}^{n} \left(2\mathbb{E}\left[W_i^2\right] - 2\mathbb{E}\left[W_i\right]^2\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \left(\mathbb{E}\left[W_i^2\right] + \mathbb{E}\left[\tilde{W}_i^2\right] - 2\mathbb{E}\left[W_i\right]\mathbb{E}\left[\tilde{W}_i\right]\right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left[\left(W_i - \tilde{W}_i\right)^2\right] \\
&= \frac{1}{2} n \mathbb{E}\left[d(W, \tilde{W})\right] \\
&\geq \frac{1}{2} n \epsilon \left(1 - \frac{1}{N_s}\right)
\end{aligned}$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \leq \binom{n}{m+1} V$$

$$\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J = s\right) \geq \frac{1}{2} n\epsilon\left(1-\frac{1}{N_s}\right)$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \operatorname{Var}(Z_i|Z_J)\right] \le \binom{n}{m+1} V$$

$$\sum_{i\notin J} \operatorname{Var}(Z_i|Z_J = s) \ge \frac{1}{2} n\epsilon \left(1 - \frac{1}{N_s}\right)$$

Note

Z is uniformly distributed over $\mathcal{Z}$ and there are N_s columns with $z_J = s$:

$$P(Z_J = s) = \frac{N_s}{|\mathcal{Z}|}$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \le \binom{n}{m+1} V$$

$$\mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \ge \sum_{s\in\mathcal{Z}_j} \frac{N_s}{|\mathcal{Z}|}\frac{1}{2} n\epsilon \left(1-\frac{1}{N_s}\right)$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \leq \binom{n}{m+1} V$$

$$\sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \geq \sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \sum_{s\in\mathcal{Z}_j} \frac{N_s}{|\mathcal{Z}|}\frac{1}{2}n\epsilon\left(1-\frac{1}{N_s}\right)$$

Inequality

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \operatorname{Var}(Z_i | Z_J)\right] \leq \binom{n}{m+1} V$$

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \operatorname{Var}(Z_i | Z_J)\right] \geq \binom{n}{m} \frac{1}{2} n\epsilon - \frac{1}{2} n\epsilon \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \sum_{s \in \mathcal{Z}_j} \frac{1}{|\mathcal{Z}|}$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \leq \binom{n}{m+1} V$$

$$\sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \geq \binom{n}{m}\frac{1}{2}n\epsilon - \frac{1}{2}n\epsilon \sum_{\substack{J\subseteq\{1,\dots,n\}\\|J|=m}} \frac{|\mathcal{Z}_J|}{|\mathcal{Z}|}$$

Inequality

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J\right)\right] \leq \binom{n}{m+1} V$$

$$\sum_{\substack{J \subseteq \{1,\ldots,n\} \\ |J|=m}} \mathbb{E}\left[\sum_{i \notin J} \mathrm{Var}\left(Z_i | Z_J\right)\right] \geq \binom{n}{m} \frac{1}{2} n\epsilon - \frac{1}{2} n\epsilon \binom{n}{m} \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}$$

Inequality

$$\sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \leq \binom{n}{m+1} V$$

$$\sum_{\substack{J\subseteq\{1,\ldots,n\}\\|J|=m}} \mathbb{E}\left[\sum_{i\notin J} \mathrm{Var}\left(Z_i|Z_J\right)\right] \geq \binom{n}{m}\frac{1}{2}n\epsilon\left(1-\frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right)$$

$$\binom{n}{m}\frac{1}{2}n\epsilon\left(1-\frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right)\leq\binom{n}{m+1}V$$

$$\binom{n}{m}\frac{1}{2}n\epsilon\left(1-\frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right)\leq\binom{n}{m+1}V$$

Remember

$$N\left(\epsilon,\mathcal{C},L_1(Q)\right)\leq D\left(\epsilon,\mathcal{C},L_1(Q)\right)\leq|\mathcal{Z}|$$

$$|\mathcal{Z}| \leq \frac{\overline{|\mathcal{Z}_J|} n\epsilon(m+1)}{n\epsilon(m+1) - 2nV + 2mV} \leq \frac{\overline{|\mathcal{Z}_J|}\epsilon(m+1)}{\epsilon(m+1) - 2V}$$

$$|\mathcal{Z}| \leq \frac{\overline{|\mathcal{Z}_J|} n\epsilon(m+1)}{n\epsilon(m+1) - 2nV + 2mV} \leq \frac{\overline{|\mathcal{Z}_J|}\epsilon m}{\epsilon m - 2V}$$

Proof

Corollary 2.6.3

For a *VC*-class of sets of index $V(\mathcal{C})$, one has

$$\max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j} \leq \left(\frac{n \cdot e}{V(\mathcal{C}) - 1} \right)^{V(\mathcal{C})-1} .$$

Proof

Corollary 2.6.3

For a *VC*-class of sets of index $V(\mathcal{C})$, one has

$$\max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j} \leq \left(\frac{n \cdot e}{V(\mathcal{C}) - 1} \right)^{V(\mathcal{C})-1} .$$

Note

$\mathcal{Z}_I$ corresponds to the collection of $C \cap \{x_i : i \in I\}$

Proof

Corollary 2.6.3

For a *VC*-class of sets of index $V(\mathcal{C})$, one has

$$\max_{x_1,\ldots,x_n} \Delta_n(\mathcal{C}, x_1, \ldots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j} \leq \left(\frac{n \cdot e}{V(\mathcal{C}) - 1} \right)^{V(\mathcal{C})-1} .$$

Note

$\mathcal{Z}_I$ corresponds to the collection of $C \cap \{x_i : i \in I\}$

Consequence

$$|\mathcal{Z}| \leq \frac{\overline{|\mathcal{Z}_J|}\epsilon m}{\epsilon m - 2V} \leq \frac{\sum_{j=0}^{V} \binom{m}{j} \epsilon m}{\epsilon m - 2V} \leq \left(\frac{e}{V} \right)^V \frac{m^{V+1}\epsilon}{m\epsilon - 2V}$$

Proof

Optimatization

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq |\mathcal{Z}| \leq \left(\frac{e}{V}\right)^V \frac{m^{V+1}\epsilon}{m\epsilon - 2V}$$

Proof

Optimatization

$$N\left(\epsilon,\mathcal{C},L_1(Q)\right) \leq |\mathcal{Z}| \leq \left(\frac{e}{V}\right)^V \frac{m^{V+1}\epsilon}{m\epsilon - 2V}$$

Optimal m

Differentiating gives $m = 2(V+1)/\epsilon$ as the optimal solution.

Proof

Optimatization

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq |\mathcal{Z}| \leq \left(\frac{e}{V}\right)^V \frac{m^{V+1}\epsilon}{m\epsilon - 2V}$$

Optimal m

Differentiating gives $m = 2(V+1)/\epsilon$ as the optimal solution.

Note

- Discretizing m changes the upper bound only by a constant factor
- $V \leq m < n$
- n depends on Q, but we can make this arbitrarily large

Proof

Optimatization

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq |\mathcal{Z}| \leq \left(\frac{e}{V}\right)^V \frac{m^{V+1}\epsilon}{m\epsilon - 2V}$$

Optimal m

Differentiating gives $m = 2(V+1)/\epsilon$ as the optimal solution.

Note

- Discretizing m changes the upper bound only by a constant factor
- $V \leq m < n$
- n depends on Q, but we can make this arbitrarily large

Bound

$$N(\epsilon, \mathcal{C}, L_1(Q)) \leq K(V+1)(4e)^V \left(\frac{1}{\epsilon}\right)^V$$

Lemma 2.6.6

Lemma 2.6.6

Let Z be an arbitrary random vector taking values in $\mathcal{Z} \subset \{0,1\}^n$ that corresponds to a VC-class $\mathcal{C}$ of subsets of a set of points $\{x_1, \ldots, x_n\}$. Then

$$\sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$$

Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$

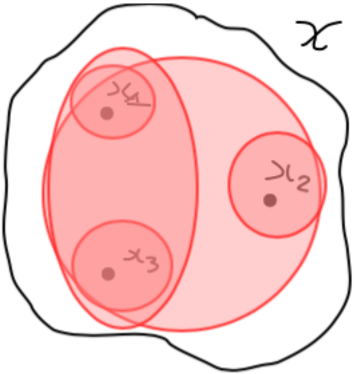
Probabilities

$Z_i | Z_j, j \neq i$ means that Z can only have two values. Call these v and w, where $v_i = 0$ and $w_i = 1$. Write $p(z) := P(Z = z)$, then $Z_i = 1$ or 0 with probabilities

$$p := \frac{p(w)}{p(w) + p(v)} \qquad 1 - p = \frac{p(v)}{p(w) + p(v)}$$

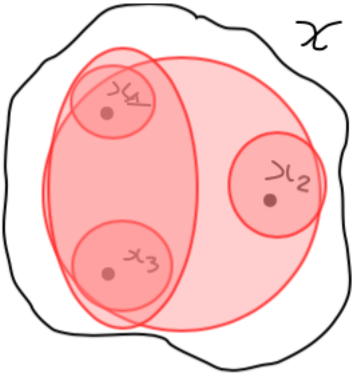
Proof idea

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$$



Proof idea

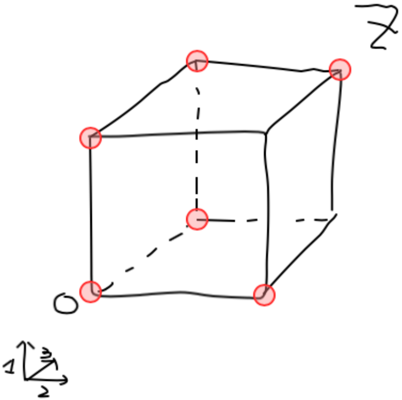
$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$



$$\begin{pmatrix}1\\0\\0\end{pmatrix} \quad \begin{pmatrix}1\\0\\1\end{pmatrix} \quad \begin{pmatrix}1\\1\\1\end{pmatrix} \quad \begin{pmatrix}0\\1\\0\end{pmatrix} \quad \begin{pmatrix}0\\0\\1\end{pmatrix} \quad \begin{pmatrix}0\\0\\0\end{pmatrix}$$

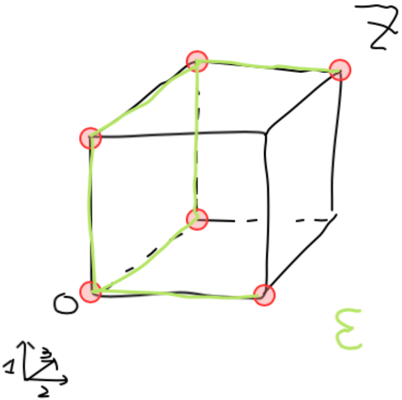
Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$



Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$

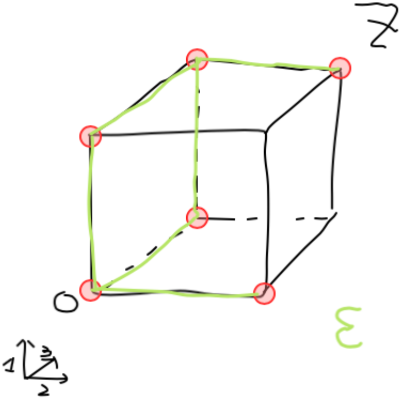


Notation

Let $\mathcal{E}$ be all edges in the graph and $\mathcal{E}_i$ be the set of edges that cross the ith dimension.

Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$

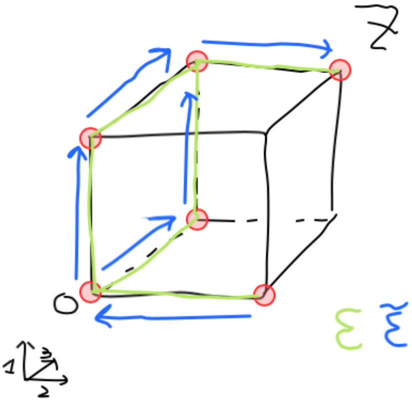


$$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] = \sum_{i=1}^{n} \sum_{\{v,w\}\in\mathcal{E}_i} (p(v) + p(w)) \cdot p(1-p)$$

$$= \sum_{\{v,w\}\in\mathcal{E}} p(v) \cdot p(w)$$

Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$



$$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] = \sum_{\{v,w\} \in \mathcal{E}} p(v) \cdot p(w)$$

$$= \sum_{v \in \mathcal{Z}} \sum_{(v,w) \in \tilde{\mathcal{E}}} p(v) \cdot p(w)$$

Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$

$$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] = \sum_{\{v,w\} \in \mathcal{E}} p(v) \cdot p(w)$$
$$= \sum_{v \in \mathcal{Z}} \sum_{(v,w) \in \tilde{\mathcal{E}}} p(v) \cdot p(w)$$

Note

Problem 2.6.5 shows that $\tilde{\mathcal{E}}$ can be formed such that

$$\sum_{(v,w) \in \tilde{\mathcal{E}}} 1 \leq V(\mathcal{C}) - 1, \quad v \in \mathcal{Z}$$

Proof idea

$\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] \leq V(\mathcal{C}) - 1$

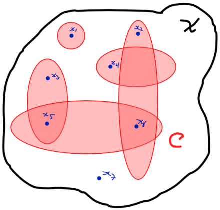
$$
\begin{aligned}
\sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Var}\left(Z_i | Z_j, j \neq i\right)\right] &= \sum_{\{v,w\} \in \mathcal{E}} p(v) \cdot p(w) \\
&= \sum_{v \in \mathcal{Z}} \sum_{(v,w) \in \tilde{\mathcal{E}}} p(v) \cdot p(w) \\
&\leq \sum_{v \in \mathcal{Z}} p(v) \cdot (V(\mathcal{C}) - 1) \\
&= V(\mathcal{C}) - 1
\end{aligned}
$$

VC-Classes of Functions

Theorem 2.6.4

There exists a universal constant K such that for any VC-class $\mathcal{C}$ of sets, any probability measure Q, any $r \geq 1$, and $0 < \epsilon < 1$,

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

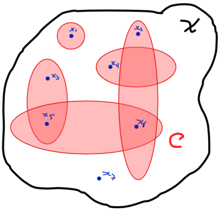


VC-Classes of Functions

Theorem 2.6.4

There exists a universal constant K such that for any VC-class $\mathcal{C}$ of sets, any probability measure Q, any $r \geq 1$, and $0 < \epsilon < 1$,

$$N(\epsilon, \mathcal{C}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$



Question

Can we also find such a result for function classes?

Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

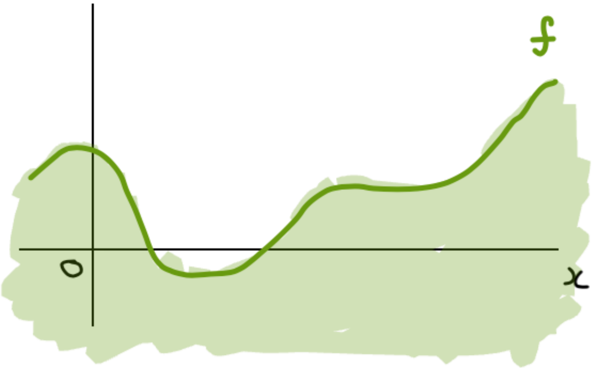
$$\{(x, t) : t < f(x)\}$$

Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\}$$



Subgraphs

Subgraph

The *subgraph* of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\}$$

VC-Class

A collection $\mathcal{F}$ of measurable functions is called a *VC-class*, if the collection of all subgraphs of the functions in $\mathcal{F}$ forms a VC-class of sets in $\mathcal{X} \times \mathbb{R}$.

Main Theorem

Theorem 2.6.7

For a VC-class of functions with measurable envelope function F and $r \geq 1$, one has for any probability measure Q with $\|F\|_{Q,r} > 0$,

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)},$$

for a universal constant K and $0 < \epsilon < 1$.

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Collection of sets

Let $\mathcal{C}$ be the collection of subgraphs C_f of functions $f \in \mathcal{F}$.

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Collection of sets

Let $\mathcal{C}$ be the collection of subgraphs C_f of functions $f \in \mathcal{F}$.

Goal

$$N\left(\epsilon 2QF, \mathcal{F}, L_1(Q)\right) \to N\left(\epsilon, \mathcal{C}, L_1(P)\right)$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

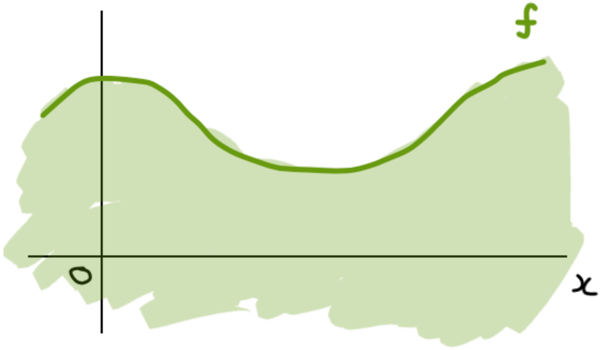
We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

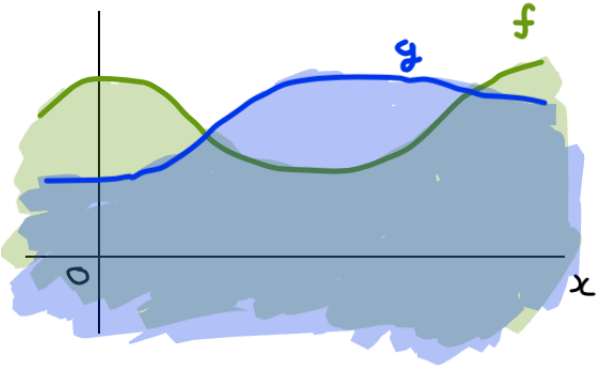


Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

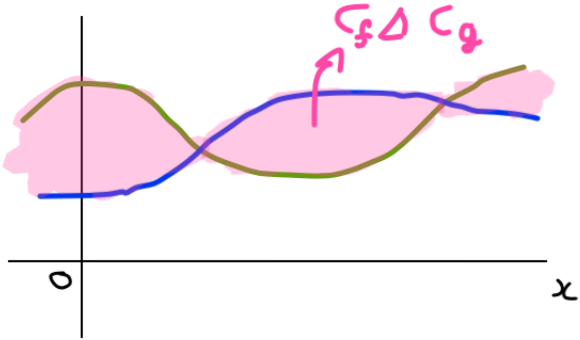


Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

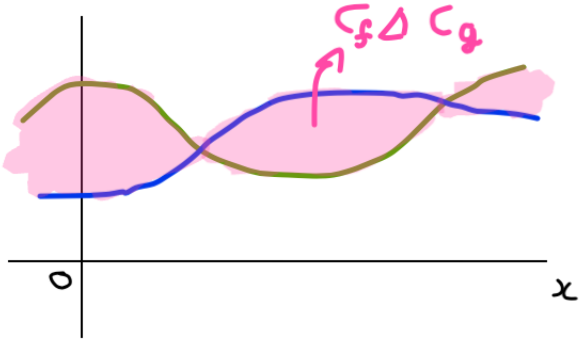


Proof

$$N\left(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q)\right)\leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q\times\lambda(C_f\Delta C_g)=Q(|f-g|)$.



$$f(x)\wedge g(x)\leq t\leq f(x)\wedge g(x)+|f(x)-g(x)|,\quad x\in\mathcal{X}$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

$$Q \times \lambda(C_f \Delta C_g) = \int_{\mathcal{X}} \int_{\mathbb{R}} \mathbb{1}_{\{t: f(x) \wedge g(x) \leq t \leq f(x) \wedge g(x) + |f(x)-g(x)|\}} d\lambda(t) dQ(x)$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Transform distances

We want to show $Q \times \lambda(C_f \Delta C_g) = Q(|f-g|)$.

$$\begin{aligned}
Q \times \lambda(C_f \Delta C_g) &= \int_{\mathcal{X}} \int_{\mathbb{R}} \mathbb{1}_{\{t: f(x)\wedge g(x) \leq t \leq f(x)\wedge g(x) + |f(x)-g(x)|\}} d\lambda(t) dQ(x) \\
&= \int_{\mathcal{X}} |f(x) - g(x)| dQ(x) \\
&= Q(|f-g|)
\end{aligned}$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Note

F is an envelope function so $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$.

Proof

$$N\left(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Note

F is an envelope function so $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$.

Measure

Renormalize $Q \times \lambda$ to a probability measure on the set $\{(x,t) : |t| \leq F(x)\}$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Note

F is an envelope function so $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$.

Measure

Renormalize $Q \times \lambda$ to a probability measure on the set $\{(x,t) : |t| \leq F(x)\}$

Total mass

$$Q \times \lambda(\{(x,t) : |t| \leq F(x)\}) = 2Q(F)$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Note

F is an envelope function so $|f(x)| \leq F(x)$ for all $x \in \mathcal{X}$.

Measure

Renormalize $Q \times \lambda$ to a probability measure on the set $\{(x,t) : |t| \leq F(x)\}$

Total mass

$$Q \times \lambda(\{(x,t) : |t| \leq F(x)\}) = 2Q(F)$$

So $P := \frac{Q\times\lambda}{2Q(F)}$ is a probability measure on $\{(x,t) : |t| \leq F(x)\}$.

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Covering numbers

$$Q(|f-g|) = Q \times \lambda(C_f \Delta C_g) = 2Q(F) \cdot P(C_f \Delta C_g)$$

Proof

$$N\left(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q)\right)\le KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\tfrac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Covering numbers

$$Q(|f-g|)=Q\times\lambda(C_f\Delta C_g)=2Q(F)\cdot P(C_f\Delta C_g)$$

Conclusion

$$N\left(\epsilon QF,\mathcal{F},L_1(Q)\right)=N\left(\epsilon/2,\mathcal{C},L_1(P)\right)\le KV(\mathcal{F})\left(\frac{8e}{\epsilon}\right)^{V(\mathcal{F})-1}$$

Proof

$$N\left(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q)\right)\le KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Covering numbers

$$Q(|f-g|)=Q\times\lambda(C_f\Delta C_g)=2Q(F)\cdot P(C_f\Delta C_g)$$

Conclusion

$$N\left(\epsilon QF,\mathcal{F},L_1(Q)\right)=N\left(\epsilon/2,\mathcal{C},L_1(P)\right)\le KV(\mathcal{F})\left(\frac{8e}{\epsilon}\right)^{V(\mathcal{F})-1}$$

Note

$$\epsilon\|F\|_{Q,1}=\epsilon Q(|F|)\ge\epsilon Q(F)$$

Proof

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \le KV(\mathcal{F})(16e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}$$

Covering numbers

$$Q(|f-g|) = Q \times \lambda(C_f \Delta C_g) = 2Q(F) \cdot P(C_f \Delta C_g)$$

Conclusion

$$N\left(\epsilon QF, \mathcal{F}, L_1(Q)\right) = N\left(\epsilon/2, \mathcal{C}, L_1(P)\right) \le KV(\mathcal{F})\left(\frac{8e}{\epsilon}\right)^{V(\mathcal{F})-1}$$

Note

$$\epsilon\|F\|_{Q,1} = \epsilon Q(|F|) \ge \epsilon Q(F)$$

Rest of the proof

For $r > 1$, us a measure R with density $\frac{F^{r-1}}{QF^{r-1}}$ such that

$$Q|f-g|^r \le 2^{r-1} R|f-g| QF^{r-1}$$