WIKIPEDIA

Covering number

In mathematics, a **covering number** is the number of spherical balls of a given size needed to completely cover a given space, with possible overlaps. Two related concepts are the *packing number*, the number of disjoint balls that fit in a space, and the *metric entropy*, the number of points that fit in a space when constrained to lie at some fixed minimum distance apart.

Contents

Definition

Let (M, d) be a metric space, let K be a subset of M, and let r be a positive real number. Let $B_r(x)$ denote the ball of radius r centered at x. A subset C of M is an *r-external covering* of K if:

$$K \subseteq \cup_{x \in C} B_r(x).$$

In other words, for every $y \in K$ there exists $x \in C$ such that $d(x, y) \leq r$.

If furthermore C is a subset of K, then it is an *r-internal covering*.

The **external covering number** of K, denoted $N_r^{\text{ext}}(K)$, is the minimum cardinality of any external covering of K. The **internal covering number**, denoted $N_r^{\text{int}}(K)$, is the minimum cardinality of any internal covering.

A subset P of K is a *packing* if $P \subseteq K$ and the set $\{B_r(x)\}_{x \in P}$ is pairwise disjoint. The **packing number** of K, denoted $N_r^{\text{pack}}(K)$, is the maximum cardinality of any packing of K.

A subset S of K is *r-separated* if each pair of points x and y in S satisfies $d(x, y) \geq r$. The **metric entropy** of K, denoted $N_r^{\text{met}}(K)$, is the maximum cardinality of any r-separated subset of K.

Examples

1. The metric space is the real line $\mathbb{R}$. $K \subset \mathbb{R}$ is a set of real numbers whose absolute value is at most k. Then, there is an external covering of $\lceil 2k/r \rceil$ intervals of length r, covering the interval $[k, -k]$. Hence:

$$N_r^{\text{ext}}(K) \leq (2k/r)$$

2. The metric space is the Euclidean space $\mathbb{R}^m$ with the Euclidean metric. $K \subset \mathbb{R}^m$ is a set of vectors whose length (norm) is at most k. If K lies in a d-dimensional subspace of $\mathbb{R}^m$, then:[1]:337

$$N_r^{\text{ext}}(K) \leq (2k\sqrt{d}/r)^d.$$

3. The metric space is the space of real-valued functions, with the l-infinity metric. The covering number $N_r^{\text{int}}(K)$ is the smallest number k such that, there exist $h_1, \ldots, h_k \in K$ such that, for all $h \in K$ there exists $i \in \{1, \ldots, k\}$ such that the supremum distance between h and h_i is at most r. The above bound is not relevant since the space is ∞-dimensional. However, when K is a compact set, every covering of it has a finite sub-covering, so $N_r^{\text{int}}(K)$ is finite.[2]:61

Properties

1. The internal and external covering numbers, the packing number, and the metric entropy are all closely related. The following chain of inequalities holds for any subset *K* of a metric space and any positive real number *r*.[3]

$$N_{2r}^{\text{met}}(K) \leq N_r^{\text{pack}}(K) \leq N_r^{\text{ext}}(K) \leq N_r^{\text{int}}(K) \leq N_r^{\text{met}}(K)$$

2. Each function except the internal covering number is non-increasing in *r* and non-decreasing in *K*. The internal covering number is monotone in *r* but not necessarily in *K*.

The following properties relate to covering numbers in the standard Euclidean space $\mathbb{R}^m$:[1]:338

3. If all vectors in K are translated by a constant vector $k_0 \in \mathbb{R}^m$, then the covering number does not change.

4. If all vectors in K are multiplied by a scalar $k \in \mathbb{R}$, then:

for all r: $N_{|k| \cdot r}^{\text{ext}}(k \cdot K) = N_r^{\text{ext}}(K)$

5. If all vectors in K are operated by a Lipschitz function ϕ with Lipschitz constant k, then:

for all r: $N_{|k| \cdot r}^{\text{ext}}(\phi \circ K) \leq N_r^{\text{ext}}(K)$

Application to machine learning

Let K be a space of real-valued functions, with the l-infinity metric (see example 3 above). Suppose all functions in K are bounded by a real constant M. Then, the covering number can be used to bound the generalization error of learning functions from K, relative to the squared loss:[2]:61

$$\text{Prob}\left[\sup_{h \in K} |\text{GeneralizationError}(h) - \text{EmpiricalError}(h)| \geq \epsilon\right] \leq N_r^{\text{int}}(K) \cdot 2\exp\frac{-m\epsilon^2}{2M^4}$$

where $r = \dfrac{\epsilon}{8M}$ and m is the number of samples.

See also

- Polygon covering

- Kissing number

References

1. Shalev-Shwartz, Shai; Ben-David, Shai (2014). *Understanding Machine Learning – from Theory to Algorithms*. Cambridge University Press. ISBN 9781107057135.
2. Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). *Foundations of Machine Learning*. USA, Massachusetts: MIT Press. ISBN 9780262018258.
3. Tao, Terrance. "Metric entropy analogues of sum set theory" (http://terrytao.wordpress.com/2014/03/19/metric-entropy-analogues-of-sum-set-theory/). Retrieved 2 June 2014.