Lecture 23: Sampling in Geometric Range Spaces

11/23/2025

1 Set Systems and Range Spaces

Set systems arise in many applications. A set system consists of a pair (P, R) where P is a set and R is a collection of subsets of P. When P is finite, we have a finite set system. Sometimes finite set systems are thought of as hypergraphs with P as vertices and each $r \in R$ as a hyperedge.

Here we will be concerned with set systems that arise in geometric settings where P is typically all of $\mathbb{R}^d$ for some dimension d, or a finite subset of $\mathbb{R}^d$. We also consider R to be sets that are induced by structured shapes that intersect with P.

1.1 Examples of Shapes

Examples of shapes include:

- Intervals
- Disks
- Half-spaces
- Polygons

In the geometric setting, (P, R) are often called **range spaces** and each $r \in R$ is called a **range**. Typically we associate r with a shape such as an interval I, and then

$$r = I \cap P$$

Geometric range spaces have additional properties that lead to a number of applications, and the notion of VC dimension, ε-sample theorem, ε-net theorem and others have had striking influence on many areas, in particular machine learning where the notion of VC dimension arose.

2 VC Dimension

VC dimension of a set system is one important measure of the complexity of a set system.

Definition 1. Let (P, R) be a range space. A finite subset $Q \subseteq P$ is said to be **shattered** by R if $\{Q \cap r : r \in R\} = 2^Q$. In other words, $R|_Q = 2^Q$, the powerset of Q.

Example 1. Suppose P is the real line and R is the collection of all closed intervals.

It can be seen from the figure that $Q = \{a, b\}$ can be shattered by the collection of intervals:

- $\emptyset$
- $\{a\}$
- $\{b\}$
- $\{a, b\}$

Definition 2. The **VC dimension** of a set system (P, R) is the maximum cardinality of a finite set $Q \subseteq P$ such that Q is shattered by R.

Example 2. Let $P = \mathbb{R}$ and R be the collection of intervals. Then VC-dim $= 2$.

Why? We saw that it is at least 2. Can it be ≥ 3?

Suppose $Q = \{a, b, c\}$ where $a < b < c$. Can we get the set $\{a, c\}$ as an intersection of $\{a, b, c\}$ and an interval? No.

Example 3. $P = \mathbb{R}^2$ (the 2D plane) and $R = \{D : D \text{ is a closed disk in the plane}\}$.

VC dimension is 3. Three points can be shattered but not 4.

Example 4. $P = \mathbb{R}^d$ and $R =$ set of half-spaces.

Recall: a half-space is defined by an inequality $\sum a_i x_i \geq b$ for some $a_1, a_2, \ldots, a_d, b \in \mathbb{R}$.

Claim: VC-dim $= d + 1$.

It is easy to see that VC-dim $\geq d + 1$. Take the $d + 1$ points: $(0, 0, \ldots, 0)$ and $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, $\ldots$, $(0, 0, \ldots, 0, 1)$.

This set can be shattered. Why?

However, $d + 2$ points cannot be shattered, and this follows from Radon's theorem.

Theorem 1 (Radon's Theorem)**.** *Let Q be a set of $d + 2$ points in $\mathbb{R}^d$. Then one can partition Q into S_1 and S_2 such that*

$$\textit{convex-hull}(S_1) \cap \textit{convex-hull}(S_2) \neq \emptyset$$

The preceding theorem shows Q cannot be shattered by half-spaces.

3 Sauer's Lemma

Now that we have seen the definition of VC dimension, we state and prove a key technical lemma about set systems with bounded VC dimension.

Theorem 2 (Sauer's Lemma)**.** *Suppose a set system (P, R) has VC dimension at most d. Let $Q \subseteq P$ be a finite set of cardinality n. Then*

$$|R|_Q| \leq \sum_{i=0}^{d} \binom{n}{i} \leq n^d$$

Proof. By induction on n.

If $n = 0$, it is trivial.

Let Q be a set of n points, $n > 0$. We can restrict attention to $R|_Q = \{r \cap Q : r \in R\}$. Hence we can work with a finite range space. Now all we need to do is count $|R|_Q|$.

Fix some $p \in Q$.

Let $R_1 = \{r \setminus \{p\} : r \in R\}$ be the set of all ranges obtained by removing p from the original ranges.

Suppose r is such that $p \in r$ and also $r \setminus \{p\} \in R$. Then both r and $r \setminus \{p\}$ project to the same range in R_1. So to count $|R|_Q|$ we create a separate range space.

Let $R_2 = \{r \setminus \{p\} : r \cup \{p\} \in R \text{ and } r \setminus \{p\} \in R\}$.

From this explanation we have:

Claim 3.

$$|R|_Q| = |R_1| + |R_2|$$

Now we consider the two range spaces $(Q \setminus \{p\}, R_1)$ and $(Q \setminus \{p\}, R_2)$.

Claim 4. *VC-dim$(Q \setminus \{p\}, R_1) \leq d$.*

Proof. Removing a point does not increase VC-dim. □

Claim 5. *VC-dim of $(Q \setminus \{p\}, R_2) \leq d - 1$.*

Proof. If $Q' \subseteq Q \setminus \{p\}$ is shattered by R_2, then since every range $r \in R_2$ satisfies the property that $r \cup \{p\}$ and $r \setminus \{p\} \in R$, we would have $Q' \cup \{p\}$ is shattered by R. Thus $|Q'| \leq d - 1$. □

Now by induction:

$$|R_1| \le \sum_{i=0}^{d} \binom{n-1}{i}$$

and

$$|R_2| \le \sum_{i=0}^{d-1} \binom{n-1}{i}$$

Thus,

$$\begin{aligned} |R|_Q| &= |R_1| + |R_2| \\ &\le \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d-1} \binom{n-1}{i} \\ &= \sum_{i=0}^{d} \binom{n}{i} \end{aligned}$$

□

3.1 Shattering Dimension

In many settings, the only way VC-dim is used is via the bound given by Sauer's lemma. So it makes sense to define the following:

Definition 3. The **shattering dimension** of a range space (P, R) is d if $\forall Q \subseteq P$ with $|Q| = n$, the size of $R|_Q \le O(n^d)$.

VC-dim$(P, R) = d \Rightarrow$ shattering-dim$(P, R) = d$.
Converse is also true with weaker parameters:
Shattering-dim$(P, R) = d \Rightarrow$ VC-dim$(P, R) = O(d \log d)$.

3.2 Closure Properties

One important aspect of VC-dim is a kind of closure when combining range spaces.

Theorem 6. *Suppose (P, R_1) and (P, R_2) are range spaces with VC-dim d_1 and d_2 respectively. Then:*

- *VC-dim of (P, R) where $R = \{r_1 \cap r_2 : r_1 \in R_1, r_2 \in R_2\}$ is $O(d_1 d_2)$.*
- *Similarly for (P, R) where $R = \{r_1 \cup r_2 : r_1 \in R_1, r_2 \in R_2\}$.*

4 ε-Sample and ε-Net Theorem

We now discuss two theorems about how a random sample of a set from a set system (P, R) can approximate it.

For the following discussion, it is useful to think of P as a finite set. Some of the concepts can be lifted to infinite sets with appropriate generalizations.

For a given system (P, R), let $\mu(r)$ denote the measure of r, i.e., $\mu(r) = \frac{|r|}{|P|}$.

Suppose we take a small random sample Q from P. Does Q preserve the measure of all $r \in R$?

For this, define $\mu_Q(r) = \frac{|r \cap Q|}{|Q|}$.

Clearly a small sample cannot touch all ranges, so we need to allow some additive error.

Definition 4. A subset $Q \subseteq P$ is an **ε-sample** for (P, R) if

$$|\mu_Q(r) - \mu(r)| \le \varepsilon \quad \forall r \in R$$

A related notion is the following:

Definition 5. Q is an ε**-net** for (P, R) if $|Q \cap r| \geq 1$ for all $r \in R$ where $\mu(r) \geq \varepsilon$.

Note that an ε-sample is automatically an ε-net.

Theorem 7. *Let (P, R) be a range space with VC dimension d. Let $\ell = \frac{cd}{\varepsilon^2} \log \frac{d}{\delta}$.*
Then a random sample of ℓ points with repetition from P is an ε-sample with probability $\geq 1 - \delta$.

Note: The sample size does not depend on $|P|$. Could be infinite.
Note: The theorem relies only on the growth rate of the number of distinct ranges of a given size that follows from Sauer's lemma. Hence the proof is not so tied to VC dimension itself.
A stronger bound is known for ε-nets:

Theorem 8. *Let (P, R) be a range space with VC-dim $\leq d$. Let $\ell = \frac{cd}{\varepsilon} \log \frac{1}{\delta}$.*
Then a random sample of ℓ points with repetition from P is an ε-net with probability $\geq 1 - \delta$.

5 Proof of ε-Sample Theorem

It is a clever argument using a double sample argument.
First we recall the additive Chernoff bound:

Theorem 9 (Chernoff Bound)**.** *Let $X_1, X_2, \ldots, X_n \in \{0, 1\}$ and independent. Let $Y = \sum_{i=1}^n X_i$ and $\mu = \mathbb{E}[Y]$. Then*

$$\Pr[|Y - \mu| \geq t] \leq 2e^{-\frac{2t^2}{n}}$$

Equivalently, $\Pr[|Y - \mu| \geq \varepsilon n] \leq 2e^{-2\varepsilon^2 n}$.

To see how to use the above theorem in our setting: Fix a range r with $\mu(r) = p \in [0, 1]$. Suppose we take an ℓ-sample Q with repetition. What is $\mathbb{E}[\mu_Q(r)]$?
Let X_i be indicator random variable for sample i being in r.
Let $Y = \sum_{i=1}^{\ell} X_i$. Then $\mu_Q(r) = \frac{Y}{\ell}$ and $\mathbb{E}[Y] = \mu(r) \cdot \ell$.
Hence $\mathbb{E}[\mu_Q(r)] = \mu(r)$.
Therefore, $\Pr[|\mu_Q(r) - \mu(r)| \geq \varepsilon] \leq 2e^{-2\varepsilon^2 \ell}$.

5.1 Full Proof of ε-Sample Theorem

It is a clever argument using a double sample argument.
Let Q be a sample of size ℓ. Let Q' be an independent sample also of size ℓ.
Let A be the event that $\exists$ some range r such that $|\mu_Q(r) - \mu(r)| > \varepsilon$. Here $\mu(r) = \frac{|r|}{|P|}$ and $\mu_Q(r) = \frac{|r \cap Q|}{|Q|}$ for short.
Let B be the event that $\exists r$ such that $|\mu_Q(r) - \mu_{Q'}(r)| > 2\varepsilon$.

Claim 10. $\Pr[A] \leq 2 \Pr[B]$.

Proof. Let D be the event that $\exists r$ such that $|\mu_Q(r) - \mu(r)| > \varepsilon$ and $|\mu_{Q'}(r) - \mu(r)| \leq \varepsilon$.
We have:

$$\begin{aligned} \Pr[B] &= \Pr[D] + \Pr[D \text{ and } A] \\ &\geq \Pr[D|A] \cdot \Pr[A] \end{aligned}$$

We claim that $\Pr[D|A] \geq \frac{1}{2}$, which would imply that $\Pr[A] \leq 2 \Pr[B]$.
To see this, suppose event A happens: $\exists r$ such that $|\mu_Q(r) - \mu(r)| > \varepsilon$. For this r, via the additive Chernoff bound,

$$\Pr[|\mu_{Q'}(r) - \mu(r)| > \varepsilon] \leq 2e^{-2\varepsilon^2 \ell} \leq \frac{1}{2}$$

This is because r is fixed and the sample is big enough and Q' is independent of Q.

If $|\mu_Q(r) - \mu(r)| > \varepsilon$ and $|\mu_{Q'}(r) - \mu(r)| \le \varepsilon$, then by triangle inequality:

$$\begin{aligned} |\mu_Q(r) - \mu_{Q'}(r)| &\ge |\mu_Q(r) - \mu(r)| - |\mu_{Q'}(r) - \mu(r)| \\ &> \varepsilon - \varepsilon = 0 \end{aligned}$$

In fact, $|\mu_Q(r) - \mu_{Q'}(r)| > 2\varepsilon$ (by considering the case when $\mu_Q(r) - \mu(r) > \varepsilon$ and $|\mu_{Q'}(r) - \mu(r)| \le \varepsilon$, so $\mu_Q(r) - \mu_{Q'}(r) \ge (\mu(r) + \varepsilon) - (\mu(r) + \varepsilon) = 2\varepsilon$ in the worst case).

Thus $\Pr[D|A] \ge \frac{1}{2}$. □

Thus we can focus on event B, which is $\Pr[\exists r : |\mu_Q(r) - \mu_{Q'}(r)| > 2\varepsilon]$ for some range r. Within factor of 2 we will get $\Pr[A]$.

To analyze B, we think of the process differently. Instead of picking Q and Q' independently as two separate steps, we think of picking 2ℓ elements Q_0 and then splitting Q_0 into two halves Q and Q'.

$$\begin{aligned} \Pr[B] &= \sum_{Q_0} \Pr[Q_0] \cdot \Pr[B|Q_0] \\ &\le \max_{Q_0} \Pr[B|Q_0] \end{aligned}$$

What is $\max_{Q_0} \Pr[B|Q_0]$?

We think of Q_0 as an arbitrary multiset of 2ℓ points and Q and Q' are obtained by evenly splitting of Q_0 into ℓ points each.

What is the advantage of this? If we fix Q_0, then $R|_{Q_0}$ has $\le (2\ell)^d$ ranges.

Thus we need to only worry about a small number of ranges.

For any fixed range in $R|_{Q_0}$, via the additive Chernoff bound:

$$\Pr[|\mu_Q(r) - \mu_{Q_0}(r)| > \varepsilon] \le 2e^{-2\varepsilon^2 \ell}$$

$$\Pr[|\mu_{Q'}(r) - \mu_{Q_0}(r)| > \varepsilon] \le 2e^{-2\varepsilon^2 \ell}$$

This is because $\ell = \frac{cd}{\varepsilon^2} \log \frac{d}{\delta}$.

Since $|\mu_Q(r) - \mu_{Q'}(r)| \le |\mu_Q(r) - \mu_{Q_0}(r)| + |\mu_{Q_0}(r) - \mu_{Q'}(r)|$:

$$\Pr[|\mu_Q(r) - \mu_{Q'}(r)| > 2\varepsilon] \le 4e^{-2\varepsilon^2 \ell}$$

By the union bound:

$$\begin{aligned} \Pr[B] &\le (2\ell)^d \cdot 4e^{-2\varepsilon^2 \ell} \\ &\le \delta \end{aligned}$$

Therefore $\Pr[A] \le 2\delta$, which gives us $\Pr[A] \le 1 - \delta$ for appropriate choice of constants.