Lecture 18: VC Dimension

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

Data Mining and Machine Learning
August–December 2020

Representational capacity

PAC learning guarantee

Let $\mathcal{H}$ be a hypothesis class, $\delta, \epsilon > 0$ and S a training set of size $n \geq \frac{1}{\epsilon}(\ln|\mathcal{H}| + \ln(1/\delta))$ drawn using D. With probability $\geq 1-\delta$, every $h \in \mathcal{H}$ with true error $\text{err}_D > \epsilon$ has training error $\text{err}_S > 0$.

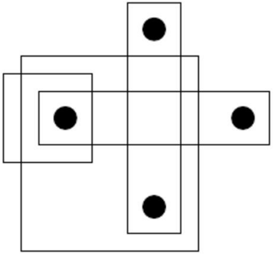
Uniform convergence

Let $\mathcal{H}$ be a hypothesis class, $\delta, \epsilon > 0$. If a training set S of size $n \geq \frac{1}{2\epsilon^2}(\ln|\mathcal{H}| + \ln(2/\delta))$ is drawn using D, then with probability $\geq 1-\delta$, every $h \in \mathcal{H}$ satisfies $|\text{err}_S(h) - \text{err}_D(h)| \leq \epsilon$.

- $|\mathcal{H}|$ is representational capacity, when $\mathcal{H}$ is finite
- How do we adapt and apply these bounds when $\mathcal{H}$ is infinite?

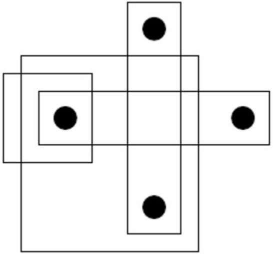
Shattering

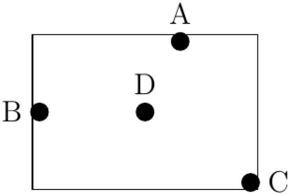
- Set system: $(X, \mathcal{H})$
  - X is a set — instance space
  - $\mathcal{H}$, set of subsets of X — set of possible classifiers / hypotheses
- $A \subseteq X$ is shattered by $\mathcal{H}$ if every subset of A is given by $A \cap h$ for some $h \in \mathcal{H}$
  - Every way of splitting A is captured by a hypothesis in $\mathcal{H}$
  - $2^{|A|}$ different subsets of A
- Example:
  - $X = \mathbb{R} \times \mathbb{R}$
  - $\mathcal{H}$: Axis-parallel rectangles
  - A : Four points forming a diamond
  - $\mathcal{H}$ shatters A



VC-Dimension [Vapnik-Chervonenkis]

- VC-Dimension of $\mathcal{H}$ — size of the largest subset of X shattered by $\mathcal{H}$
  - For axis-parallel rectangles, VC-dimension is at least 4
  - Not a universal requirement — some sets of size 4 may not be shattered
- No set of size 5 can be shattered by axis-parallel rectangles
  - Draw a bounding box rectangle — each edge touches a boundary point
  - At least one point lies inside the bounding box
  - Any set that includes the boundary points also includes the interior point





VC-Dimension, Examples

- Intervals of reals have VC-dimension 2
  - $X = \mathbb{R}$, $\mathcal{H} = \{[a, b] \mid a \leq b \in \mathbb{R}\}$
  - Cannot shatter 3 points: consider subset with first and third point
- Pairs of intervals of reals have VC-dimension 4
  - $X = \mathbb{R}$, $\mathcal{H} = \{[a, b] \cup [c, d] \mid a \leq b, c \leq d \in \mathbb{R}\}$
  - Cannot shatter 5 points: consider subset with first, third and fifth point
- Finite sets of real numbers
  - $X = \mathbb{R}$, $\mathcal{H} = \{Z \mid Z \subseteq \mathbb{R}, |Z| < \infty\}$
  - Can shatter any finite set of reals — VC-dimension is infinite
- Convex polygons, $X = \mathbb{R} \times \mathbb{R}$
  - For any n, place n points on unit circle
  - Each subset of these points is a convex polygon — VC-dimension is infinite

VC-Dimension, Examples

Half spaces in d dimensions

- $X = \mathbb{R}^d$, classifiers are linear separators, $\{x \mid w^T x \leq t\}$
- Can shatter $d+1$ points consisting of origin and d unit vectors
- For a subset D, set $w_i = -1$ if unit vector i in D, $w_i = 1$ if i not in D
- $w^T x \leq -1$ characterizes D, if origin not in D
  - If C contains the origin, use $w^T x \leq 0$

Theorem (Radon)

Any set $S \subseteq \mathbb{R}^d$ with $|S| \geq d+2$, can be partitioned into two disjoint subsets A and B such that $\text{convex-hull}(A) \cap \text{convex-hull}(B) \neq \emptyset$.

- Hence no set of size $d+2$ is shattered by half-spaces in $\mathbb{R}^d$

VC-Dimension, Examples

Spheres in d dimensions

- $X = \mathbb{R}^d$, sphere with centre x_0 and radius r, $\{x \mid |x - x_0| \leq r\}$
- Cannot shatter any set S with $d{+}2$ points
  - Suppose we could. Consider A, B given by Radon's theorem
  - Spheres C_A and C_B such that $C_A \cap S = A$, $C_B \cap S = B$
  - C_A and C_B may intersect, but intersection is disjoint from S
  - Separating halfplane perpendicular to line joining centres — contradiction
- Consider $d{+}1$ points consisting of d unit vectors plus the origin
  - Let D be a subset of these $d{+}1$ points, with $|D| = k$
  - Centre x_0 is sum of vectors in D
  - For $\left\{\begin{array}{ll} v \in D, & |v - x_0| = \sqrt{k-1}, \\ v \notin D, & |v - x_0| = \sqrt{k+1}, \\ \text{origin}, & |0 - x_0| = \sqrt{k} \end{array}\right]$ — choose radius r appropriately

Summary

- PAC learning and uniform convergence use size of finite hypothesis set as measure of representational capacity
- VC-dimension provides a way of measuring capacity for infinite hypothesis sets
- VC-dimension may be finite or infinite
- We will see how to apply VC-dimension to generalize theorems relating sample size, true error and training error