CS 446: Machine Learning
Discussion Session

Daniel Khashabi

October 16, 2015

Reminder: Recall that the VC dimension of a class C is the largest number d such that there exists a set S of size d such that any partitioning of the set S into $+$ and $-$ is valid (this is the definition of shattering).

1 VC dimension

- $\mathcal{H}$ is axis parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^2$.
  VC Dimension 4
- $\mathcal{H}$ is axis-parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^3$.
  VC Dimension 6
  Hint: first show that it can shatter the 6 points (1,0,0),(0,1,0),(0,0,1),(1,0,0),(0,1,0),(0,0,1). If we draw a bounding box for these points, then by excluding/including each point by moving a face of the box, we can get any labeling for the points. For 7 points, consider the bounding box. If the bounding rectangle has at least one point in its interior, then we cannot accomplish the labeling where the interior point is labeled - and the rest are labeled +.
- $\mathcal{H}$ is axis-parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^d$.
  VC Dimension $2d$
- $\mathcal{H}$ is the union of 2 intervals, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 4.
- $\mathcal{H}$ is $1\{a \sin(x) > 0\}$, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 1.
- $\mathcal{H}$ is $1\{\sin(x + a) > 0\}$, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 2.

2 Concept class and VC-dimension:

Show that a finite concept class C has VC dimension at most $\log |C|$.
Hint: The different number of ways d points can be labeled with $\pm$ is 2^d.

Reminder: Define:

- Concept class C defined over an instance space X, containing n instances.
- A learner L using a hypothesis space H.

C is PAC learnable by L using H if

- For all $f \in C$
- For all distributions D over X

with probability at least $(1-\delta)$ the learner L produces a hypothesis with error at most ϵ, given m instances from X (where m is polynomial function of $1/\epsilon$, $1/\delta$, $|H|$, n).

Note: "Efficiently Learnable", if the learner produces the output in polynomial time of $1/\epsilon$, $1/\delta$, $|H|$, n.

Sample complexities:

- Consistent learning:
$$m > \Omega\left(\frac{1}{\epsilon}\log\frac{1}{\delta} + \frac{|H|}{\epsilon}\right)$$
- Agnostic learning:
$$m > \Omega\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta} + \frac{d}{\epsilon^2}\right)$$

3 PAC-learning and VC-dimension:

Show that if a concept class C has infinite VC dimension, then it is not PAC-learnable.
Hint:
$$m > \Omega\left(\frac{1}{\epsilon}\log\frac{1}{\delta} + \frac{d}{\epsilon}\right)$$