ILLC Project Course in Statistical Learning Theory

Mathias Winther Madsen
mathias.winther@gmail.com

Institute for Logic, Language, and Computation
University of Amsterdam

January 2015

VC Dimension

Problem

A parliament consisting of 100 MPs vote on an agenda of 8 items, and they manage to all disagree with each other on at least one issues.

This implies that the agenda contained at least 4 highly divisive items that split the parliament into 16 distinct voting blocks.

Why?

VC Dimension

If an agenda $|x| = t$ contains no divisive subagendas $|z| = k$, then the agenda splits the parliament into $\leq \Phi(k, t)$ factions.

How many factions can there be for an agenda $|x| = t + 1$ if it contains no divisive subagenda $|z| = k$?

1. Create a hold-out agenda $h = x \setminus \{x_{t+1}\}$.
2. Count the number of factions.
3. Select the factions that will be divided when x_{t+1} is put to the vote, and put them in a committee.
4. If this committee is split into m factions by $z \subseteq h$, then the parliament is split into $2m$ by $z \cup \{x_{t+1}\} \subseteq x$.
5. Hence, the committee cannot be (completely) divided by any $z \subseteq h$ with $|z| = k - 1$. Hence, the committee is split into at most $\Phi(k - 1, t)$ factions by h.

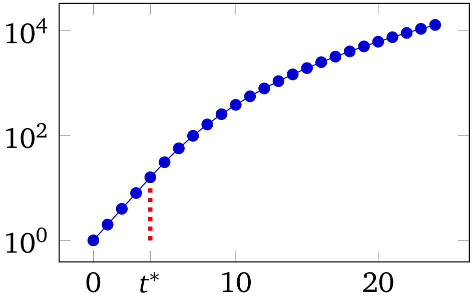
VC Dimension

Definition

A portfolio is said to **shatter** a set $x \subseteq \Omega$ if

$$N_S(t \,|\, x) \;=\; |\, S \downarrow x \,| \;=\; 2^{|x|}.$$

Its **VC dimension** is the size of the largest set it can shatter.



VC Dimension

Examples

1. The Glivenko-Cantelli sets: $S = \{\{r \leq \theta\} \mid \theta \in \mathbb{R}\}$.
2. Histogram bins: $S = \{(wz, wz + w] \mid z \in \mathbb{Z}\}$.
3. Half-planes: $S = \{y \leq ax + b \mid a, b \in \mathbb{R}^2\}$.
4. Half-spaces: $S = \{\mathbf{v} \cdot \mathbf{x} \leq \theta \mid a, b \in \mathbb{R}^2\}$.

VC Dimension

Theorem

If $t^* < \infty$, then the training set frequencies frequencies of the sets $A \in S$ converge uniformly to their test set frequencies.

Proof.

Fix an x with $|x| = 2t$, letting $S_{2t} := (S \downarrow x)$. Then, given x,

$$\begin{aligned}
\Pr\{\exists A : |f_1(A) - f_2(A)| > \varepsilon\} &\leq \sum_{A \in S_{2t}} \Pr\{|f_2(A) - f_2(A)| > \varepsilon\} \\
&\leq N(2t) \times 2\exp(-2\varepsilon^2 t).
\end{aligned}$$

This upper bound tends to 0 when $t \to \infty$ regardless of x. □