Lesson 8: VC Dimensions

Infinite hypothesis spaces

$$m \geq \frac{1}{\epsilon}(ln|H| + ln(\frac{1}{\delta}))$$

Haussler theorem breaks with infinite hypothesis spaces.

For example: linear separators, artificial NN, decision tree with continous input.

Maybe not so bad

Consider example:

$X : \{1, 2, ..., 10\}$

$H : h(x) = \{x \leq \theta \text{ where } \theta \text{ is real}\}$

Syntactic hypotheses space: All the things you can write

Semantic hypothesis space: All the functionally different hypothesis.

Though there are infinite hypotheses, there are find semantically relevant hypotheses.

Power of a hypothesis space

What is the largest set of inputs that the hypothesis class can label in all possible ways?

In the above hypothesis: $H : h(x) = \{x \leq \theta \text{ where } \theta \text{ is real}\}$, there is no way for a pair of data points to be labeled in all possible ways. It can however label **one** point is all possible ways (two ways).

What VC stands for

Shattering – labeling in all possible ways.

Vapnik-Chervonenkis dimension – the largest set of inputs shattered by a class of learners.

This is related with the smount of data needed for learning.

Quiz: interval training

X = Reals $H = \{h(x) = \{x$ in $[a, b]$ – a real interval$\}\}$

Has VC dimension of 2.

To prove:

To prove VC is at least something: There exists a arrangment of data points, for all labelling combination, there exists a hypothesis that can separate them correctly.

To prove VC is less than something: For all arrangment of data points, there exists a labelling combination, s.t. there is no hypothesis can separate them.

Quiz: linear separators

$X = R^2$

$H = \{h(x) = w^T x >= \theta\}$

Has VC dimension of 3.

The rings

The d-dimensional hyperplane as VC dim $d + 1$

VC dimension often is the number of the parameters.

Quiz: polygons (convex)

$X : \mathbb{R}^2$

H: points inside some convex polygon (any number of edges)

VC dimension is "infinite". (Put all data points on a circle, connect the the positive points to form a convex polygon).

Sample complexity and VC dimension

Update Haussler's theorem with VC dimension:

infinite case:

$$m \geq \frac{1}{\epsilon}(8 * VC(H) * lg(\frac{13}{\epsilon}) + 4 * lg(\frac{2}{\delta}))$$

finite case:

$$m \geq 1/\epsilon(ln|H| + ln(1/\delta))$$

VC dimension of finite H

Upper bound: $d = VC(H)$ implies there exist 2^d (binary classification problem) distinct concepts (each gets a different h)

$2^d \leq |H|$ and $d \leq log_2|H|$

Fundamental Theorem of Machine Learning

H is PAC-learnable if and only if VC dimension is finite.

Summary

- VC dimension. Shattering
- VC relates to hypothesis space parameters ("true" number of parameters)
- VC relates to finite hypothesis space size.
- Sample complexity related to VC dimension.
- VC dimension captures PAC-learnability