VC Dimension

A Tutorial for the Course *Computational Intelligence*

http://www.igi.tugraz.at/lehre/CI

Stefan Häusler
Institute for Theoretical Computer Science
Inffeldgasse 16b/I

Abstract

This script contains the collection of examples for the computation of the VC-dimension as presented in the lecture on the 4/4/2003. There is also a printable PDF version[1] of the script.

1 Intervals

What is the VC-dimension of intervals in $\mathbb{R}$? The target function is specified by an interval, and labels any example positive iff it lies inside that interval.

VC-dim = 2. A set of two points can be shattered, since there's only a single block of positive examples that could lie within the interval. But no set of 3 points can be shattered, because it can not be labeled in alternating $+, -, +$ order.

2 Axis Parallel Rectangles

What is the VC-dimension of axis parallel rectangles in the plane $\mathbb{R}^2$? The target function is specified by a rectangle, and labels any example positive iff it lies inside that rectangle.

VC-dim = 4. For instance the set of points $\{(1,0),(0.1),(-1,0),(0,-1)\}$ can be shattered, but if you draw the smallest enclosing box around 5 points it is not possible to label the point inside the box $-$ and the remaining points on the edges $+$.

3 Circles

What is the VC-dimension of circles in the plane $\mathbb{R}^2$? I.e., examples are points in $\mathbb{R}^2, C = \{c_{p,r} : p \in \mathbb{R}^2, r \in \mathbb{R}\}$ and $c_{p,r} = 1$ if x is within distance r of p. Or, in words, a legal target function is specified by a circle, and labels any example positive iff it lies inside that circle.

VC-dim = 3. It is easy to see the VC-dimension is at least 3 since any 3 points that make up a non-degenerate triangle can be shattered.It is a bit trickier to prove that the VC-dimension is less than 4. Given 4 points, the easy case is when one is inside the convex hull of th others. In that case, because circles are convex, it is not possible to label the inside point $-$ and the outside points $+$. Otherwise, call the points a, b, c, d in clockwise order. the claim is that it is not possible for one circle c_1

[1] VC_examples.pdf

to achieve labeling $+, -, +, -$ and for some other c_2 to achieve labeling $-, +, -, +, -$. If such c_1, c_2 existed, then their symmetric difference would consist of 4 disjoint regions, which is impossible for circles.

4 Triangles

What is the VC-dimension of triangles in the plane $\mathbb{R}^2$? I.e., a legal target function is specified by a triangle, and labels any example positive iff it lies inside that triangle.

VC-dim $= 7$. Given 7 points on a circle, they can be labeled in any desired way because in any labeling, the negative examples form at most 3 contiguous blocks. Therefore one edge of the triangle can be used to cut off each block. However, no set of 8 points can be shattered. If one of the points is inside the convex hull of the rest, then it is not possible to label that point negative and the rest positive. Otherwise, it is not possible to label them in alternating $+, -, +, -, +, -, +, -$ order.

5 Halfspaces in $\mathbb{R}^n$

Prove that the VC-dimension of the class H_n of halhspaces in n dimensions is $n+1$. (H_n is the set of functions $w_1x_1 + ... + w_nx_n \leq w_0$, where $w_0, ..., w_n$ are real-valued.) We will use the following definition: the *convex hull* of a set of points S is the set of all convex combinations of points S; this is the set of all points that can be written as $\sum_{x_i \in S} \lambda_i x_i$, where $\lambda_i \leq 0$, and $\sum_i \lambda_i = 1$. It is not hard to see that if a halfspace has all points from a set S on one side, then the entire convex hull of S must be on that side as well.

5.1 Lower Bound

Prove that VC-dim(H_n)$\geq n+1$ by presenting a set of $n+1$ points in n-dimensional space such that one can partition that set with halfspaces in all possible ways. (And, show how one can partition the set in any desires way.)

One good set of $n+1$ points is: The origin and all points with a 1 in one coordinate and zeros in the rest (i.e., all neighbors of the origin on the Boolean cube). Let p_i be the point with a 1 in the ith coordinate. Suppose we wish to partition this set into two pieces S_1 and S_2 with a hyperplane (and, say the origin is in S_1). Then just choose the hyperplane:

$$\sum_{\{i: p_i \in S_2\}} x_i = 1/2. \tag{1}$$

5.2 Upper Bound Part I

The following is "Radon's Theorem".

Theorem. *Let S be a set of $n+2$ points in n dimensions. Then S can be partitioned into two (disjoint) subsets S_1 and S_2 whose convex hulls intersect.*

Show that Radon's Theorem implies that the VC-dimension of halfspaces is *at most* $n+1$. Conclude that VC-dim$(H_n) = n+1$.

If S is a set of $n+2$ points, then by Radon's theorem we may partition S into sets

S_1 and S_2 whose convex hulls intersect. Let $p \in S_1$ be a point in that intersection. Assume there exist a hyperplane

$$w \cdot x_i \leq w_0, \qquad \forall x_i \in S_1$$
$$w \cdot x_i > w_0, \qquad \forall x_i \in S_2$$

so that

$$w \cdot p \leq w_0$$

which is contradicted by

$$w \cdot p = \sum_{i:x_i \in S_2} \lambda_i w \cdot x_i > (\sum_{i:x_i \in S_2} \lambda_i) \min_{i:x_i \in S_2} (w \cdot x_i) = \min_{i:x_i \in S_2} (w \cdot x_i) > w_0.$$

So no set of $n+2$ points can be shattered and VC-dim(H_n) $= n+1$.

5.3 Upper Bound Part II

Prove Radon's Theorem. As a first step show the following. For a set of $n+2$ points $x_1, ..., x_{n+2}$ in n-dimensional space, there exist $\lambda_1, ..., \lambda_{n+2}$ *not all zero* such that $\sum_i \lambda_i x_i = 0$ and $\sum_i \lambda_i = 0$. (This is called *affine dependence.*)

We first prove the affine dependence.Let $v_1, ..., v_{n+2}$ be defined as $v_i = \langle x_i, 1 \rangle \in \mathbb{R}^{n+1}$. So $\{v_i\}$ is linearly dependent. And so there exist a set of scalares $\lambda_1, ..., \lambda_{n+2}$, not all zero, so that $\sum_i \lambda_i v_i = 0$. These $\{\lambda_i\}$ satisfy the required conditions.

Further $S_1 = \{x_i | \lambda_i \geq 0\}$ and $S_2 = \{x_i | \lambda_i < 0\}$. So

$$\lambda^* = \sum_{i:x_i \in S_1} \lambda_i = -\sum_{j:x_j \in S_2} |\lambda_j|.$$

Define

$$x^* = \sum_{i:x_i \in S_1} \lambda_i x_i = -\sum_{j:x_j \in S_2} |\lambda_j| x_j.$$

Consider the point

$$\frac{x^*}{\lambda^*} = \sum_{i:x_i \in S_1} \frac{\lambda_i}{\lambda^*} x_i = -\sum_{j:x_j \in S_2} \frac{|\lambda_j|}{\lambda^*} x_j.$$

This point lies in the convex hull of both S_1 and S_2.