Gautam Ramasubramanian
Rutgers DIMACS Research

VC Dimension

Definitions

Suppose we has a class of hypothesis functions H.

Shattering:
If $\exists$ a positioning of n training examples, such that $\forall$ possible labelings of those n points, $\exists$ $h \in H$ such that h perfectly separates the training data, then the H is said to *shatter* n points.

VC Dimension:
VC Dimension is a property of H and it is the largest number of points H can shatter.

Example

1. H consists of hyperplanes that can separate the data. (What this means is that H consists of functions whose decision boundary is a hyperplane). What is the VC dimension?

   The VC dimension will be 3, because for the "general" positioning of the 3 points, there exists a hypothesis that can separate them no matter what label they assume.

   However, every positioning of 4 points has a labeling that cannot be separated by any hyperplane. For example,

2. Consider H to be the set of all circles in $\mathbb{R}^2$, such that all the points inside the circle are labeled 1 and all the points outside the circle are labeled 0. What is the VC dimension of H?

   The VC Dimension is still 3, for the same reasons as above. Here are pictures that show.

Graduate Problem

$\mathcal{F}$ is a vector space of functions with dimension r. Say that each function in $\mathcal{F}$ goes from $\mathbb{R}^d$ to $\mathbb{R}$. The set A_f is defined as follows.

$$A_f = x : f(x) \geq 0$$

Let H, the set of all hypotheses, be defined as such.

$$H = A_f : f \in F$$

H is a class of subsets of $\mathbb{R}^d$, where data points x live. Prove that VC dimension of H is less than or equal to r. Equivalently, prove that H cannot shatter $r+1$ points in $\mathbb{R}^d$.

Solution:

Its a bit confusing because the hypotheses in H are actually $sgn(f)$, where $f \in \mathcal{F}$, but the decision boundaries of those functions are represented through the $H = A_f : f \in F$ definition. The goal is to prove that for any positioning of $r+1$ points in $\mathbb{R}^d$, there exists a labelling for those points such that no hypothesis in H can separate them. Suppose we chose arbitrary $r+1$ points $x_1, x_2, \ldots, x_{r+1}$. An A_f would separate them if $\forall i = 1 \ldots r+1$

$$label(x_i) = sgn(f(x_i))$$

We can say that f produces the labeling *label*. We have to prove that there exists a labeling for $x_1, x_2, \ldots, x_{r+1}$ that no function in $\mathcal{F}$ can produce.

Let us consider the mapping $g : \mathcal{F} \leftarrow \mathbb{R}^{r+1}$ such that

$$g(f) = (f(x_1), f(x_2), \ldots, f(x_{r+1})) = b$$

The range of g will be denoted as $\mathcal{B}$, so $g(f_i) = b_i$. Note that g is a linear mapping, so that means $\mathcal{B}$ is a vector space of dimension d that is spanned by $b_1, b_2, \ldots, b_{r+1}$. The following is an important property.

Property: If b^s and b^t are orthogonal to each other and nonzero, then they cannot produce the same labelling for $x_1, x_2, \ldots, x_{r+1}$. ($f^s$ and f^t are just 2 different functions in $\mathcal{F}$, nothing to do with exponentiation).

Proof: If f^s and f^t are orthogonal, then their inner product must be 0. That means that b^s and b^t have an inner product of 0.

$$b^s = (f^s(x_1), f^s(x_2), \ldots, f^s(x_{r+1}))$$

$$b^t = (f^t(x_1), f^s(x_2), \ldots, f^t(x_{r+1}))$$

$$b^s \cdot b^t = f^s(x_1)f^t(x_1) + f^s(x_2)f^t(x_2) + \ldots + f^s(x_{r+1})f^t(x_{r+1}) = 0$$

However, if f^s and f^t produce the same labelling for $x_1, x_2, \ldots, x_{r+1}$, then

$$\forall i = 1, 2, \ldots, r+1$$

$$sgn(f^s(x_i)) = sgn(f^t(x_i))$$

which means

$$\forall i = 1, 2, \ldots, r+1$$

$$f^s(x_i)f^t(x_i) > 0 \text{ since both } b_s \text{ and } b_t \text{ are nonzero}$$

which means that

$$f^s(x_1)f^t(x_1) + f^s(x_2)f^t(x_2) + \ldots + f^s(x_{r+1})f^t(x_{r+1}) > 0$$

which is a contradiction.

Now that the property is proven, then we can proceed to prove that $x_1, x_2, \ldots, x_{r+1}$ is not shatterable by H, or equivalently, we can find a label that we are not able to produce using any function in $\mathcal{F}$. If we look at $g(\mathcal{F}) = \mathcal{B}$, it is can be seen that B has a dimension of r, but the codomain of g, which is R^{r+1} has a dimension of $r+1$. Therefore, there exists a vector in R^{r+1} that is orthogonal to B. That vector, which we can call v^*, produces a labeling if we take the sgn function at every coordinate of v^*. This labelling cannot be produced by any function in $\mathcal{F}$, because v^* in orthogonal to every vector in $\mathcal{B}$. Therefore, there exists a labeling of points $x_1, x_2, \ldots, x_{r+1}$ that cannot be separated by any hypothesis in H, which means that the VC dimension of H is r or less.