Notes on VC Dimension of Sin/Cos Function

BY XING CHAO
xingnix@live.com

1 Problem

Suppose there is a series of scalar x_i , which can be assigned to arbitrary values according to VC dimension definition. The problem is how to choose these values to meat the all kinds of dichotomy with sine or cosine funtion?

Consider three point, $x_1 = 0, x_2 = 2, x_3 = 3$, and the corrysponding category label $c(x_i) \in \{0, 1\}$. It is eany to draw a sine/cosine function cure to shatter the three ponts:

$$
\begin{aligned}
c(x_1) = 1, c(x_2) = 0, c(x_3) = 0 \;&\rightarrow\; \cos\left(2\pi\frac{x}{5}\right) - 1 + \Delta \\
c(x_1) = 1, c(x_2) = 1, c(x_3) = 0 \;&\rightarrow\; \cos\left(2\pi\frac{x}{2}\right) - 1 + \Delta \\
c(x_1) = 1, c(x_2) = 0, c(x_3) = 1 \;&\rightarrow\; \cos\left(2\pi\frac{x}{3}\right) - 1 + \Delta
\end{aligned}
$$

where Δ is a small number.

2 Basic Idea

In order to extend the previous example of three points to more points, let's choose the value of three point again to see if there is any clue. If we can choose period of cosine function to be exactly divisible by some $x_i \in \{x_i | c(x_i) = 1\}$, and not exactly divisible by other $x_i \in \{x_i | c(x_i) = 0\}$, the cosine function can be used to classify the points according to $c(x_i)$. It is possible to shatter these points if we choose x_i carefully to make such period exist for evry category label $c(x_i)$.

Let

$$
\begin{aligned}
x_1 &= 2 \times 7 \times 13 \\
x_2 &= 3 \times 7 \times 11 \\
x_3 &= 5 \times 11 \times 13
\end{aligned}
$$

then

$$
\begin{aligned}
c(x_1) = 1, c(x_2) = 0, c(x_3) = 0 \;&\rightarrow\; \cos\left(2\pi\frac{x}{2}\right) - 1 + \Delta \\
c(x_1) = 1, c(x_2) = 1, c(x_3) = 0 \;&\rightarrow\; \cos\left(2\pi\frac{x}{7}\right) - 1 + \Delta \\
c(x_1) = 1, c(x_2) = 0, c(x_3) = 1 \;&\rightarrow\; \cos\left(2\pi\frac{x}{13}\right) - 1 + \Delta
\end{aligned}
$$

3 Algorithm

Each x_i is a production of prime numbers and for each subset $\{x_i | c(x_i) = 1\}$ of $\{x_i\}$,there is a prime number in each $x_i \in \{x_i | c(x_i) = 1\}$ and not in $x_i \notin \{x_i | c(x_i) = 1\}$. That prime number is the period of the cosine function.

Now it is time to find a way to generate x_i. First let's count the amount of prime numbers we need. Each subset of $\{x_i\}$ need one prime number and there is 2^N subset, where $N = |\{x_i\}|$ is the number of elements in $\{x_i\}$. The number "2" is a clue that binary number may give some help. A dichotomy can be represented by $c(x_1)c(x_2)\cdots c(x_n)$, where each $c(x_i)$ is 0 or 1 and it can be used to represent the index of the dichotomy. Then a corresponding prime is choosen to put in x_i whose $c(x_i) = 1$ in the binary number.

Example

$$
\begin{aligned}
c(x_1) = 1, c(x_2) = 0, c(x_3) = 0 \;\rightarrow\; & 001 \\
& x_1 = \mathrm{prod}([\mathrm{prime}(1)\,]) = \mathrm{prod}([2]) \\
& x_2 = [] \\
& x_3 = [] \\
c(x_1) = 0, c(x_2) = 1, c(x_3) = 0 \;\rightarrow\; & 010 \\
& x_1 = \mathrm{prod}([2\,]) \\
& x_2 = \mathrm{prod}([\mathrm{prime}(2)\,]) = \mathrm{prod}([3]) \\
& x_3 = [] \\
c(x_1) = 1, c(x_2) = 1, c(x_3) = 0 \;\rightarrow\; & 011 \\
& x_1 = \mathrm{prod}([2, \mathrm{prime}(3)]) = \mathrm{prod}([2, 5]) \\
& x_2 = \mathrm{prod}([3, \mathrm{prime}(3)]) = \mathrm{prod}([3, 5]) \\
& x_3 = [] \\
c(x_1) = 0, c(x_2) = 0, c(x_3) = 1 \;\rightarrow\; & 100 \\
& x_1 = \mathrm{prod}([2, 5]) \\
& x_2 = \mathrm{prod}([3, 5]) \\
& x_3 = \mathrm{prod}([\mathrm{prime}(4)\,]) = \mathrm{prod}([7]) \\
c(x_1) = 1, c(x_2) = 0, c(x_3) = 1 \;\rightarrow\; & 101 \\
& x_1 = \mathrm{prod}([2, 5, \mathrm{prime}(5)]) = \mathrm{prod}([2, 5, 11]) \\
& x_2 = \mathrm{prod}([3, 5]) \\
& x_3 = \mathrm{prod}([7, \mathrm{prime}(5)\,]) = \mathrm{prod}([7, 11]) \\
c(x_1) = 0, c(x_2) = 1, c(x_3) = 1 \;\rightarrow\; & 101 \\
& x_1 = \mathrm{prod}([2, 5, 11]) \\
& x_2 = \mathrm{prod}([3, 5, \mathrm{prime}(7)]) = \mathrm{prod}([3, 5, 13]) \\
& x_3 = \mathrm{prod}([7, 11, \mathrm{prime}(7)\,]) = \mathrm{prod}([7, 11, 13])
\end{aligned}
$$

then

$$
\begin{aligned}
x_1 &= 2 \times 5 \times 11 \\
x_2 &= 3 \times 5 \times 13 \\
x_3 &= 7 \times 11 \times 13
\end{aligned}
$$

For each dichotomy a period can be choosen according to the binary number of $c(x_1)c(x_2)\cdots c(x_n)$

$$\mathrm{period} = \mathrm{prime}(c(x_1)c(x_2)\cdots c(x_n))$$