CS771 Assignment 1

Introduction To Machine Learning

Submitted by

Jigyashu Garg (200478)
Manish Meena (200560)
Saurav Kumar (200906)
Harsh Saroha (200419)
Deepesh Pratap (200313)

February 15, 2023

PROBLEM 1

By giving a detailed mathematical derivation (as given in the lecture slides), show how a simple XORRO PUF can be broken by a single linear model. Recall that the simple XORRO PUF has just two XORROs and has no select bits and no multiplexers (see above figure and discussion on Simple XORRO PUF). Thus, the challenge to a simple XORRO PUF has just R bits. More specifically, give derivations for a map ϕ : {0,1}$^{\mathbb{R}}$ mapping R-bit 0/1-valued challenge vectors to D-dimensional feature vectors (for some D > 0) and show that for any simple XORRO PUF, there exists a linear model i.e. w $\mathbb{R}^D$, b ϵ $\mathbb{R}$ such that for all challenges c ϵ $\{0,1\}^R$, the following expression

$$\frac{1 + sign(w^T \phi(c) + b)}{2}$$

gives the correct response.

Answer : Let b_i be the output of $(i-1)^{th}$ XOR and a_i be the configuration bit of i^{th} XOR

$$b_{i+1} = (1 - a_i)b_i + a_i(1 - b_i)$$

$$b_{i+1} = (1 - 2a_i)b_i + a_i \tag{1}$$

Similarly ,

$$b_{i+1} = (1 - 2a_{i-1})b_{i-1} + a_{i-1} \tag{2}$$

Substitute 2 in 1 :
: $b_{i+1} = (1 - 2a_i)((1 - 2a_{i-1})b_{i-1} + a_{i-1}) + a_i$
: $b_{i+1} = (1 - 2a_i)(1 - 2a_{i-1})(1 - 2a_{i-2})b_{i-1} + a_{i-2}(1 - 2a_{i-1})(1 - 2a_i) + a_{i-1}(1 - 2a_i) + a_i$
: $b_{i+1} = C_{i+1}b_0 + \frac{1-C_{i+1}}{2}$

where ,

$$C_i = \prod_{k=0}^{i-1}(1 - 2a_k)$$

Now Total time :
$T = \sum_{i=0}^{R-1}\{(1-a_i)\{(1-b_i)\delta_0^i + b_i\delta_{10}^i\} + a_i\{(1-b_i)\delta_{01}^i + b_i\delta_{11}^i\}\}$
T = $t_0 + t_1$
where t_0= time when $b_0 = 0$
t_1 = time when $b_0 = 1$
$b_i^0 + b_i^1 = 1$
Therefore :
T = $t_0 + t_1 = \sum_{i=0}^{R-1}\{(1-a_i)(\delta_0^i + \delta_{10}^i) + a_i(\delta_{01}^i + \delta_{11}^i)\}$
: $= \sum_{i=0}^{R-1}\{(a_i(\delta_{01}^i + \delta_{11}^i - \delta_{10}^i - \delta_{10}^i) + (\delta_{00}^i + \delta_{10}^i)\}$
$T = \frac{1}{f} = \mathrm{w}^T\mathrm{x} + \mathrm{b}$
Where : $x_i = a_i$
: $w_i = \{\ \delta_{01}^i + \delta_{11}^i - \delta_{10}^i - \delta_{10}^i\}$
: $b = \sum_{i=0}^{R-1}(\delta_{00}^i + \delta_{10}^i)$
Now, given that if $f_1 > f_2$ output =1 and if $f_1 < f_2$ output =0
Let T_1 be the total time for 1st XORRO PUF and T_2 be the time for 2nd XORRO PUF
Let $w^T = w_2^T$ - w_1^T and B = b_2 - b_1
Now , if $f_1 > f_2$ then $T_1 < T_2$ and output =1 .
=> T_2 - $T_1 > 0$
: (w_2^Tx + b_2) - (w_1^Tx + b_1) > 0
=> (w_2^T - w_1^T)x +(b_2 - b_1) > 0
=> w^Tx + B > 0
Similarly for w^Tx + B < 0 output = 0
$\therefore$ It can be modelled as

$$\frac{1 + sign(w^T\phi(c) + b)}{2}$$

PROBLEM 2

Show how to extend the above linear model to crack an Advanced XORRO PUF. Do this by treating an advanced XORRO PUF as a collection of multiple simple XORRO PUFs. For example, you may use $\mathbf{M = 2^{S-1}(2^S - 1)}$ linear models, one for each pair of XORROs,to crack the advanced XORRO PUF.

Answer : The linear model obtained above can be extended to crack an advanced XORRO PUF by training M linear models where $M = 2^{S-1}(2^S - 1)$. Each model is uniquely defined for a pair of two XORRO, which is selected using 2S select bits provided with data. M models can be stored in a dictionary where the key is a tuple of two numbers obtained from the 1st S and 2nd S select bits respectively, and the value of the key will be the corresponding model trained for respective XORRO PUF. The corresponding model from the dictionary will be called and the prediction will be made according to the model trained from training data.