# Assignment

**Amit Kumar**
CSE-22111008
amitkm22@iitk.ac.in
IIT K

Yogesh Shrivastava
CSE-22111085
yogeshs22@iitk.ac.in
IIT K

Ayush Mohod
CSE-22111016
ayushm22@iitk.ac.in
IIT K

Anurag Kamal
CSE-22111010
akamal22@iitk.ac.in
IIT K

Ajeet Kumar
CSE-22111005
ajeetkr22@iitk.ac.in
IIT K

## Question 1

By giving a mathematical derivation, show their exists a way to map the binary digits 0, 1 to sign $-1, +1$ as say, $f : \{0, 1\} \to \{-1, +1\}$ and another way $f : \{-1, +1\} \to \{0, 1\}$ to map signs to bits ( not that m and f need to be inverses of each other) so that for any sets of binary digits $(b_1, b_2, \cdots, b_n)$ for any n $\in \mathbb{N}$, we have

$$XOR(b_1, b_2, \cdots, b_n) = f\left(\Pi_{i=1}^{n} m(b_i)\right)$$

Thus, the XOR function is not that scary – it is essentially a product.

### Solution

In this case we have a mapping function $m$ which maps

$$\begin{aligned} &0 \to 1 \\ &1 \to -1 \\ &m = 1 - 2x \text{ where } x \in \{0, 1\} \end{aligned}$$

and we have other function which satisfy reverse mapping

$$\begin{aligned} &1 \to 0 \\ &-1 \to 1 \\ &f = \frac{1-x}{2} \text{ where } y \in \{1, -1\} \end{aligned}$$

and this function m and f will satisfy the Equation

$$XOR(b_1, b_2, \cdots, b_n) = f\left(\Pi_{i=1}^{n} m(b_i)\right)$$

Let $n = 2$ i.e. we have 2 input bits $(b_1, b_2)$

**Case 1:** $b_1 = 0, b_2 = 0$
**LHS :**

$$XOR(b_1, b_2) = b_1 \oplus b_2 = 0 \oplus 0 = 0$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

**RHS :**

$f\big(\Pi_{i=1}^{2} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2)\big)$
$f(1 \times 1)$
$0$

**LHS = RHS**

**Case 2:** $b_1 = 0, b_2 = 1$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 = 0 \oplus 1 = 1$

**RHS :**

$f\big(\Pi_{i=1}^{2} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2)\big)$
$f(1 \times -1)$
$f(-1)$
$1$

**LHS = RHS**

**Case 3:** $b_1 = 1, b_2 = 0$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 = 1 \oplus 0 = 1$

**RHS :**

$\big(\Pi_{i=1}^{2} m(b_i)\big)$
$\big(m(b_1) \times m(b_2)\big)$
$f(-1 \times 1)$
$f(-1)$
$1$

**LHS = RHS**

**Case 4:** $b_1 = 1, b_2 = 1$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 = 1 \oplus 1 = 1$

**RHS :**

$f\big(\Pi_{i=1}^{2} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2)\big)$
$f(-1 \times -1)$
$f(1)$
$0$

**LHS = RHS**

Similarly for n = 3 (odd digit) $b_1, b_2, b_3$
**Case 1:** $b_1 = 0, b_2 = 0, b_3 = 1$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 \oplus b_3 = 0 \oplus 0 \oplus 1 = 1$

**RHS :**

$f\big(\Pi_{i=1}^{3} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2) \times m(b_3)\big)$
$f(1 \times 1 \times -1)$
$f(-1)$
$1$

**LHS = RHS**

**Case 2:** $b_1 = 1, b_2 = 0, b_3 = 1$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 \oplus b_3 = 1 \oplus 0 \oplus 1 = 0$

**RHS :**

$f\big(\Pi_{i=1}^{3} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2) \times m(b_3)\big)$
$f(-1 \times 1 \times -1)$
$f(1)$
$0$

**LHS = RHS**

**Case 3:** $b_1 = 1, b_2 = 1, b_3 = 1$
**LHS :**

$XOR(b_1, b_2) = b_1 \oplus b_2 \oplus b_3 = 1 \oplus 1 \oplus 1 = 0$

**RHS :**

$f\big(\Pi_{i=1}^{3} m(b_i)\big)$
$f\big(m(b_1) \times m(b_2) \times m(b_3)\big)$
$f(-1 \times -1 \times -1)$
$f(-1)$
$1$

**LHS = RHS**

Hence we see that our function m and f is working fine for every bits (either or odd) to satisfy the equation.

$$XOR(b_1, b_2, \cdots, b_n) = f\Bigg(\Pi_{i=1}^{n} m(b_i)\Bigg)$$

## Question 2

Let (u,a), (v,b), (w,c) be three linear model that can exactly predict the outputs of three individual PUF's sitting inside the XOR-PUF. For sake of simplicity, let us hide the bias term inside the model vector by adding a unit dimension to the original feature vector so that so that we have $\in \mathbb{R}^9$ The above calculations show that the response of the XOR-PUF can be easily obtained by (by applying f) if we are able to get hold of the following quantity:

$$sign(\tilde{u}^T\tilde{x}) \cdot sign(\tilde{v}^T\tilde{x}) \cdot sign(\tilde{w}^T\tilde{x})$$

To exploit the above result, first give a mathematical proof that for any real number (that could be positive, negative, zero) $r_1, r_2, \cdots r_n$ for any n $\in \mathbb{N}$, we always have

$$\Pi_{i=1}^{n} sign(r_i) = sign\big(\Pi_{i=1}^{n}\big)$$

Assume that sign(0) = 0. Make sure you address all edge cases in your calculations e.g. if one or more of the numbers is 0.

### Solution

Proof by Mathematical Induction
at n = 2 (base case)

$$\Pi_{i=1}^{2} sign(r_i) = sign(r_1) \cdot sign(r_2) \quad sign(\Pi_{i=1}^{2})r_i = sign(r_1 \cdot r_2)$$

| $r_1$ | $r_2$ | sign($r_1$) $\cdot$ $sign(r_2)$ | sign($r_1 \cdot r_2$) |
|---|---|---|---|
| 0 | +ve | sign($r_1$) $\cdot sign(r_2) = 0$ | sign($r_1 \cdot r_2) = 0$ |
| 0 | -ve | sign($r_1$) $\cdot sign(r_2) = 0$ | sign($r_1 \cdot r_2) = 0$ |
| 0 | 0 | sign($r_1$) $\cdot sign(r_2) = 0$ | sign($r_1 \cdot r_2) = 0$ |
| +ve | +ve | sign($r_1$) $\cdot sign(r_2) = +ve$ | sign($r_1 \cdot r_2) = +ve$ |
| +ve | -ve | sign($r_1$) $\cdot sign(r_2) = -ve$ | sign($r_1 \cdot r_2) = -ve$ |
| -ve | -ve | sign($r_1$) $\cdot sign(r_2) = +ve$ | sign($r_1 \cdot r_2) = +ve$ |

Assume that at $n = k \quad \Pi_{i=1}^{k} sign(r_i) = sign(\Pi_{i=1}^{k}(r_i) \cdots (1)$ holds true
For $n = k + 1$

**LHS :**

$\Pi_{i=1}^{k+1}(sign(r_i))$

$\Pi_{i=1}^{k}(sign(r_i)) \cdot sign(r_{k+1})$

From eqn (1)

$sign(\Pi_{i=1}^{k} r_i) \cdot sign(r_{k+1})$

Assume $sign(\Pi_{i=1}^{k} r_i) = x$

$sign(x) \cdot sign(r_{k+1})$

**RHS :**

$sign(\Pi_{i=1}^{k+1}(r_i))$

$sign(\Pi_{i=1}^{k} r_i \cdot r_{k+1}$

From eqn (1)

$sign(\Pi_{i=1}^{k} r_i \cdot r_{k+1})$

Since sign$(\Pi_{i=1}^{k} r_i) = x$

$sign(x \cdot r_{k+1})$

$$LHS = RHS$$

## Question 3

The above calculation tells us that we need to get hold of its following quantity

$$\tilde{u}^T\tilde{x} \cdot \tilde{v}^T\tilde{x} \cdot \tilde{w}^T\tilde{x}$$

Now show that the above can be expressed as a linear model but possibly in a different dimensional space. Show that there exists a $D$ such that $D$ depends only on the number of PUFs (in this case 3) and the dimensionality of $\tilde{x}$ ( in this case 8 + 1 = 9) and there exists a way to map 9 dimensional vector to D dimensional vectors as $\phi : \mathbb{R}^9 \to \mathbb{R}^D$ such that for any triple $(\tilde{u}, \tilde{v}, \tilde{w})$, there exists a vector $\mathbf{W} \in \mathbb{R}^D$ such that for every $\tilde{x} \in \mathbb{R}^9$ , we have $(\tilde{u}^T\tilde{x}) \cdot (\tilde{v}^T\tilde{x}) \cdot (\tilde{w}^T\tilde{x}) = \mathbf{W}^T\phi(\tilde{x})$

**Solution:**

For 3 PUF's

$$\begin{aligned}(\tilde{u}^T\tilde{x}) \cdot (\tilde{v}^T\tilde{x}) \cdot (\tilde{w}^T\tilde{x}) &= (\sum_{j=1}^{9} \tilde{u}_j\tilde{x}_j) \cdot (\sum_{j=1}^{9} \tilde{v}_j\tilde{x}_j) \cdot (\sum_{j=1}^{9} \tilde{w}_j\tilde{x}_j)) \\ &= \sum_{j=1}^{9}\sum_{k=1}^{9}\sum_{l=1}^{9} \tilde{u}_j^T\tilde{v}_k^T\tilde{w}_l^T\tilde{u}_j\tilde{v}_k\tilde{w}_l\end{aligned}$$

$\tilde{x} = (\tilde{x_1} \cdots \tilde{x_9})$ to $\phi(\tilde{x}) = (\tilde{x_1} \cdot \tilde{x_1} \cdot \tilde{x_1}, \tilde{x_1} \cdot \tilde{x_1} \cdot \tilde{x_2}, \cdots, \tilde{x_1} \cdot \tilde{x_1} \cdot \tilde{x_9}, \tilde{x_1} \cdot \tilde{x_2} \cdot \tilde{x_1} \cdots, \tilde{x_9} \cdot \tilde{x_9} \cdot \tilde{x_9})$
This is $9^3 = 729$ dimensional function

$$(\tilde{u}^T\tilde{x}) \cdot (\tilde{v}^T\tilde{x}) \cdot (\tilde{w}^T\tilde{x}) = \mathbf{W}^T\phi(\tilde{x})$$
$$\therefore W = (u_1 \cdot v_1 \cdot w_1, u_1 \cdot v_1 \cdot w_2, \cdots u_1 \cdot v_1 \cdot w_9, u_1 \cdot v_2 \cdot w_1, \cdots, u_9 \cdot v_9 \cdot w_9)$$

$\therefore$ 729 dimensions for 3 PUF's

## Question 5

For the method you implemented, describe your PDF report what were the hyperparameters e.g. step length, policy on choosing the next coordinate if doing SDCA, mini-batch size if doing MBSGD etc and how did you arrive at the best values for hyperparameters e.g. you might say *"We used step length at time t to be $\eta/\sqrt{t}$ where we checked for $\eta$ = 0.1, 0.2, 0.5, 1, 2, 5 using held out validation and found $\eta$ = 2 to work the best"*. For another example you might say, *"We tried random and cyclic coordinate selection choices and found cyclic to work best using 5-fold cross validation"* Thus, you must tell us among which hyperparameter choices did you search for the best and how.

### Solution

The hyperparameters to implement our solver are:

1. Learning rate($\eta$): whose optimal value for our model is 0.001, to find this value we used 5-fold cross validation in which at time $t$ to be $\eta/\sqrt{t}$ where we checked for random learning rate values $\eta$ = 0.1, 0.2, 0.05, 0.03, 0.002, 0.001 and found $\eta$ = 0.0001 to be giving the most optimal result for our solver.
2. Lambda parameter $\lambda$: whose optimal value for our model is 0.01, to find this value we need the same above of 5-fold cross validation method where we checked for $\lambda$ = 0 to 2 (positive value only) and found $\lambda$ = 0.01 to be giving the optimal result for our solver.

For applying 5-fold cross validation we have split the data in a way that 80% of data will be used for training and 20% of data for testing to find optimal value of learning rate($\eta$) and lambda parameter ($\lambda$).

## Question 6

Plot the convergence curves in your PDF report offered by your chosen method as we do in lecture notebooks. The x axis in the graph should be time taken and the y axis should be the test classification accuracy (i.e. higher is better). Include this graph in your PDF file submission as an image.

### Solution

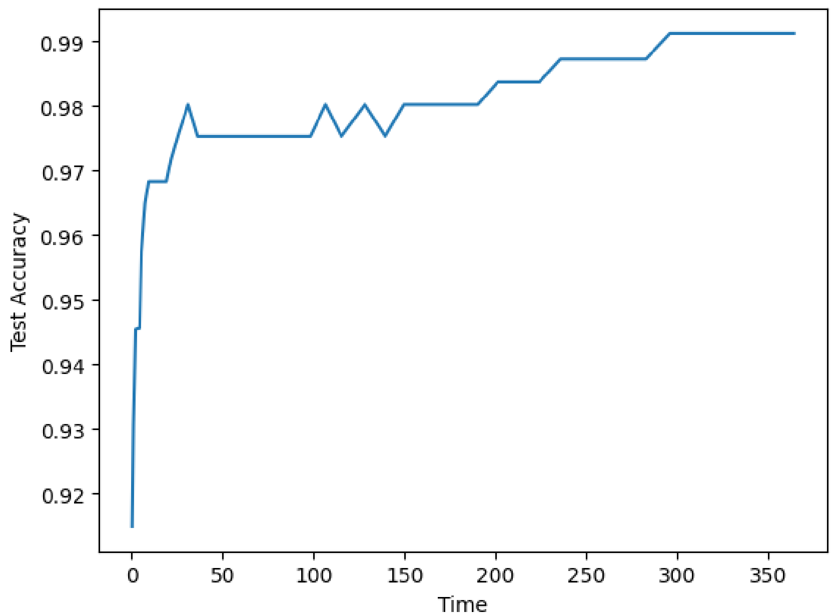


Figure 1: Convergence Curve