CS771 : ASSIGNMENT 1

Group Members
Arabinda Karmakar (22111011)
Ayush Kothiyal (22111015)
Manish Kumar Ghildiyal (22111039)
Vamshikiran Morlawar (22111066)
Vinay Agrawal (22111068)

1 Question

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

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

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

Solution:

For mapping m : {0 , 1} $\rightarrow$ {-1 , +1}, we can get a pair of values as (input , output) which are (0 , -1) and (1 , +1). Using these two value pairs, we can get an equation of line by the formula

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

where, (x_1, y_1) and (x_2, y_2) corresponds to values pairs (0 , -1) and (1 , +1) respectively.

On substituting these values in above equation, we get

$$\begin{aligned} y - (-1) &= \frac{1 - (-1)}{1 - 0}(x - 0) \\ y + 1 &= 2x \\ \therefore y &= (2x - 1) \end{aligned} \tag{1}$$

Similarly, for mapping f : {-1 , +1} $\rightarrow$ {0 , 1}, we can get a pair of values as (input , output) which are (-1 , 0) and (+1 , 1). Using these two value pairs, we can get an equation of line by the formula

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

where, (x_1, y_1) and (x_2, y_2) corresponds to values pairs (-1 , 0) and (+1 , 1) respectively.

On substituting these values in above equation, we get

$$y - (0) = \frac{1 - (0)}{1 - (-1)}(x - (-1)$$

$$y = \frac{x + 1}{2}$$

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

$$\therefore y = \frac{x+1}{2} \tag{2}$$

Therefore, by the above two equations we can say that there exists a mapping from binary digits 0, 1 and signs -1, +1, where map m : {0, 1} $\rightarrow$ {-1, +1} corresponds to the eq. 1 and map f : {-1 , +1} $\rightarrow$ {0 , 1} corresponds to the eq. 2.

2 Question

Let (u, a), (v, b), (w, c) be the three linear models that can exactly predict the outputs of the three individual PUFs 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 we have $\tilde{u}$ = [u, a], $\tilde{v}$ = [v, b], $\tilde{w}$ = [w, c], $\tilde{x}$ = [x, 1] $\in R^9$. The above calculation shows that the response of the XOR-PUF can be easily obtained (by applying f) if we are able to get hold of the following quantity:

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

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

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

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:

To Prove :

$$\prod_{i=1}^{n} sign(r_i) = sign(\prod_{i=1}^{n} r_i), \forall n \in \mathbb{N}$$

Proof :

Case 1 :

If there exists one or more 'i' for which r_i = 0, then in that case,

L.H.S :

$$\prod_{i=1}^{n} sign(r_i) = [sign(r_1).sign(r_2)....sign(r_k)....sign(r_n)]$$

$$= 0 \ (Since, sign(r_k) = 0 \ if \ r_k = 0)$$

R.H.S :

$$sign(\prod_{i=1}^{n} r_i) = sign(r_1.r_2....r_k....r_n)$$

$$= 0 \ (Since, sign(0) = 0 \ as \ r_k = 0)$$

Therefore, the equality holds when one or more numbers are 0.

Case 2:

If there does not exist any 'i' for which r_i = 0, then, by Mathematical Induction,

Basis Step:

For n = 1 ,

L.H.S :

$$\prod_{i=1}^{1} sign(r_i) = sign(r_i)$$

R.H.S :

$$sign(\prod_{i=1}^{1} r_i) = sign(r_i)$$

Hence, the above equality holds for n = 1 , i.e.

$$\prod_{i=1}^{1} sign(r_i) = sign(\prod_{i=1}^{1} r_i)$$

Inductive Hypothesis:

Now, let us assume that the equality holds for some n = k. Then we get,

$$\prod_{i=1}^{k} sign(r_i) = sign(\prod_{i=1}^{k} r_i)$$
$$\therefore [sign(r_1).sign(r_2)....sign(r_k)] = sign(r_1.r_2....r_k) \tag{1}$$

Inductive Step:

Now, for n = k + 1,

L.H.S :

$$\prod_{i=1}^{k+1} sign(r_i) = sign(r_1).sign(r_2)....sign(r_k).sign(r_{k+1})$$

From eq. 1:

$$[sign(r_1).sign(r_2)....sign(r_k)] = sign(r_1.r_2....r_k)$$

So,

$$\prod_{i=1}^{k+1} sign(r_i) = sign(r_1.r_2....r_k).sign(r_{k+1})$$
$$= \frac{|r_1.r_2....r_k|}{r_1.r_2....r_k} . \frac{|r_{k+1}|}{r_{k+1}}$$
$$= \frac{|r_1.r_2....r_k.r_{k+1}|}{r_1.r_2....r_k.r_{k+1}}$$
$$= sign(r_1.r_2....r_k.r_{k+1})$$
$$\therefore \prod_{i=1}^{k+1} sign(r_i) = sign(\prod_{i=1}^{k+1} r_i)$$

Hence, the equality also holds for n = k+1. Thus, by Mathematical Induction it is proved that the eqality $\prod_{i=1}^{n} sign(r_i) = sign(\prod_{i=1}^{n} r_i)\ holds\ for\ \forall n \in \mathbb{N}$

3 Question

The above calculation tells us that all we need to get hold of is the following quantity

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

Now show that the above can be expressed as a linear model but possibly in a different dimensional space. Show that there exists a dimensionality 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 vectors to D dimensional vectors as $\phi : R^9 \rightarrow R^D$ such that for any triple ($\tilde{u}$, $\tilde{v}$, $\tilde{w}$), there always exists a vector $W \in R^D$ such that for every $\tilde{x} \in R^9$, we have $(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x})$ = $W^T\phi(\tilde{x})$.

Hint: First try solving this for the simpler case where there are only 2 PUFs. If we expand the terms of $(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}) = (\sum_{j=1}^{9} \tilde{u}_j\tilde{x}_j)(\sum_{j=1}^{9} \tilde{v}_j\tilde{x}_j)$, we get an expression of the form $\sum_{j=1}^{9}\sum_{k=1}^{9} \tilde{u}_j\tilde{v}_k\tilde{x}_j\tilde{x}_k$. Thus, if we create a $9^2 = 81$-dimensional function that maps

$$\tilde{x} = (\tilde{x}_1, ..., \tilde{x}_9) \; to \; \phi(\tilde{x}) = (\tilde{x}_1\tilde{x}_1, \tilde{x}_1\tilde{x}_2, ..., \tilde{x}_1\tilde{x}_9, \tilde{x}_2\tilde{x}_1, ..., \tilde{x}_9\tilde{x}_9),$$

then we are done since we can now get $(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x})$ = $W^T\phi(\tilde{x})$. by taking

$$W = (\tilde{u}_1\tilde{v}_1, \tilde{u}_1\tilde{v}_2, ..., \tilde{u}_1\tilde{v}_9, \tilde{u}_2\tilde{v}_1, ..., \tilde{u}_9\tilde{v}_9)$$

Closely understand the trick in this simpler case and then extend it to the case of 3 PUFs to solve this part of the problem. Give detailed calculations for your solution.

Solution:

There exist a way to map 9 dimensional array to a D dimensional vectors as $\phi : R^9 \rightarrow R^D$; such that for any triple $(\tilde{u}, \tilde{v}, \tilde{w})$ there exist a vector $W \in R^9$ such that for every $\tilde{x} \in R^9$ (in this case 8+1=9)

we have,

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = W^T\phi(\tilde{x})$$

$$= (\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x})$$

$$= \sum_{i=1}^{9} \tilde{u}_i\tilde{x}_i . \sum_{j=1}^{9} \tilde{v}_j\tilde{x}_j . \sum_{k=1}^{9} \tilde{w}_k\tilde{x}_k$$

$$= \sum_{i=1}^{9}\sum_{j=1}^{9}\sum_{k=1}^{9} \tilde{u}_i.\tilde{v}_j.\tilde{w}_k.\tilde{x}_i.\tilde{x}_j.\tilde{x}_k$$

Thus , if we create a $9^3 = 729$ - dimensional function that maps

$$\tilde{x} = (\tilde{x}_1, ..., \tilde{x}_9) \; to$$

$$\phi(\tilde{x}) = (\tilde{x}_1\tilde{x}_1\tilde{x}_1, \tilde{x}_1\tilde{x}_1\tilde{x}_2, ..., \tilde{x}_1\tilde{x}_2\tilde{x}_1, \tilde{x}_1\tilde{x}_2\tilde{x}_2, ..., \tilde{x}_9\tilde{x}_9\tilde{x}_9)$$

then we get

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = W^T\phi(\tilde{x})$$

by taking

$$W = (\tilde{u}_1\tilde{v}_1\tilde{w}_1, \tilde{u}_1\tilde{v}_1\tilde{w}_2, ..., \tilde{u}_1\tilde{v}_2\tilde{w}_1, \tilde{u}_1\tilde{v}_2\tilde{w}_2, ..., \tilde{u}_9\tilde{v}_9\tilde{w}_9)$$

4 Question

Coding Question.

5 Question

For the method you implemented, describe in 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 the hyperparameters, e.g. you might say *"We used step length at time t to be $\eta/\sqrt{t}$ where we checked for η = 0.1, 0.2, 0.5, 1, 2, 5 using held out validation and found η = 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 hyper-parameter choices did you search for the best and how.

Solution:

For training the linear model W we used the stochastic gradient descent technique along with hinge loss function to reduce the overall loss incurred in each iteration. Therefore, we had to define two parameters, one for regularization and another for the step length where λ is the regularization parameter and α denotes learning rate or the step length for the convergence.

Initially, we chose $\lambda = 0$ and the learning rate $\alpha = 0.7$ but found the prediction on the test set to not be acceptable for our model due to over-fitting on the training data. So, we ran a loop for 10 iterations on n for the the new $\lambda = \frac{\lambda}{n}$ because of which λ decreases in the pattern 1, 0.3, 0.1, 0.03, 0.01 and so on.

On plotting the graph between λ and the prediction accuracy percentage we found 0.003 to be the best value for lambda. Similarly we found $0.7e^{-3}$ to be the best step length for our model when we kept decreasing the value of α by a factor of 0.1 and plotted it with respect to the cumulative loss incurred after prediction, as it provided faster convergence and an acceptable accuracy percentage for the prediction.

6 Question

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:

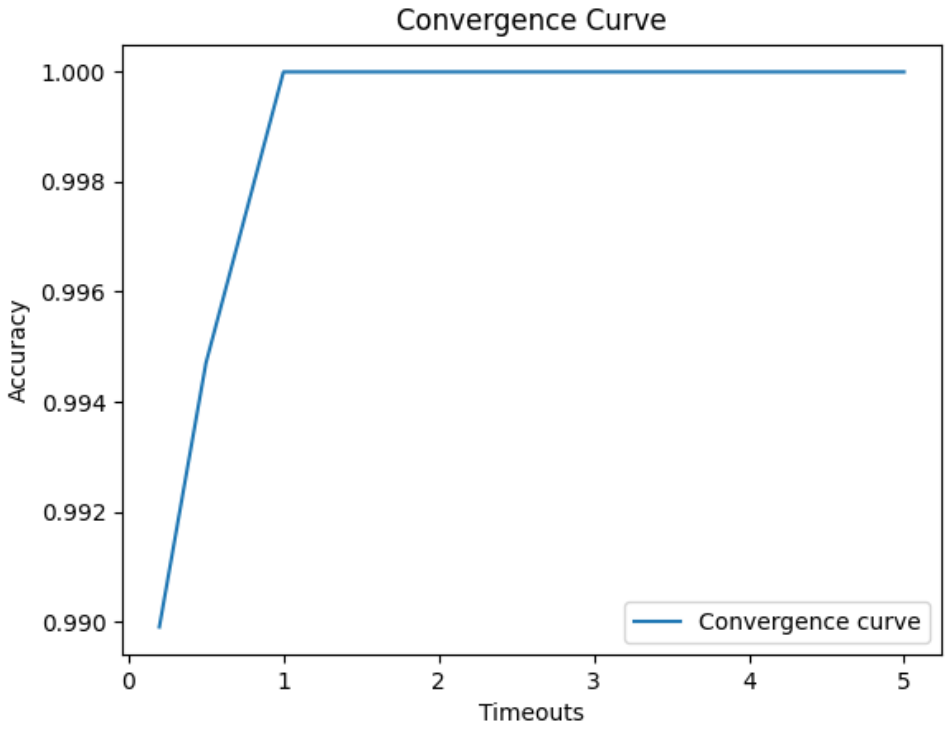

Figure 1: Convergence Curve