CS771 Introduction to Machine Learning
Assignment 1

Gunj Mehul Hundiwala
22111024
gunjmehul22@iitk.ac.in

Kartick Verma
22111029
kartickv22@iitk.ac.in

Kush Shah
22111033
kushshah22@iitk.ac.in

Raj Kumar
22111050
rajkumar22@iitk.ac.in

Saqeeb
22111053
saqeeb22@iitk.ac.in

Question 1

By giving a mathematical derivation, show there exists a way to map the binary digits 0,1 to sign -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 $n \epsilon \mathbb{N}$ we have

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

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

Solution 1

We have to find the function

$$m : \{0, 1\} \rightarrow \{-1, +1\}\ and \tag{2}$$

$$f : \{-1, +1\} \rightarrow \{0, +1\}\ so\ that \tag{3}$$

for any set of binary digit $b_1, b_2, ..., b_n$ for $n \epsilon \mathbb{N}$ we have

$$XOR(b_1, b_2, ..., b_n) = f\left(\prod_{i=1}^{n} m(b_i) \right) \tag{4}$$

let m(x) = $1 - 2x$ $x \epsilon \{0, 1\}$ and f(y) = $\dfrac{1 - y}{2}$ $y \epsilon \{-1, +1\}$

Now we will try to prove

$$XOR(b_1, b_2, ..., b_n) = f\left(\prod_{i=1}^{n} m(b_i) \right) \tag{5}$$

Using Principle of mathematical induction.

For n=2, we have 2^2

Case 1 :

$$XOR(0,0) = 0 \Rightarrow LHS$$
$$RHS = f(m(0).m(0)) = f(1.1) = 0$$
$$LHS = RHS$$

Case 2 :

$$XOR(0,1) = 1 = LHS$$
$$RHS = f(m(0).m(1)) = f(1.(-1)) = f(-1) = 1$$
$$LHS = RHS$$

Case 3 :

$$XOR(1,0) = 1 = LHS$$
$$RHS = f(m(1).m(0)) = f((-1) * (+1)) = f(-1) = 1$$
$$LHS = RHS$$

Case 4 :

$$XOR(1,1) = 0 = LHS$$
$$RHS = f(m(1).m(1)) = f((-1) * (-1)) = f(1) = 0$$
$$LHS = RHS$$

So p(2) is true

Lets assume that p(k) is true, so

$$XOR(b_1, b_2, ..., b_k) = f\left(\prod_{i=1}^{k} m(b_i)\right) \tag{6}$$

Lets check for whether p(k+1) is true or not,

$$LHS \Rightarrow XOR(b_1, b_2, ..., b_(k+1)) = XOR(b_1, ..., b_k) \oplus b_(k+1) \tag{7}$$

$$Now, XOR(b_1, ..., b_k) = f\left(\prod_{i=1}^{k} m(b_i)\right) \tag{8}$$

$\therefore$ XOR(b_1,...,b_k) can be 0 or 1 and so the quantity $f\left(\prod_{i=1}^{k} m(b_i)\right)$

Now, we will have 4 cases,

Case 1 :

$$XOR(b_1, ..., b_k) = 0, b_(k+1) = 0$$
$$LHS = XOR(b_1, ..., b_(k+1)) = 0 \oplus 0 = 0$$
$$RHS = f\left(\prod_{i=1}^{k+1} m(b_i)\right) = f\left(\prod_{i=1}^{k} m(b_i).m(b_{k+1})\right)$$
$$XOR(b_1, ..., b_k) = 0 means f\left(\prod_{i=1}^{k} m(b_i)\right) = 0$$

We know that, f() function maps $-1 \rightarrow 1, 1 \rightarrow 0$

If f() has a output 0, then input must have been 1

$\therefore \prod_{i=1}^{k} m(b_i) = 1$

Substituting in the RHS, we get

$$RHS = f(1.m(b_{k+1})) = f(m(b_{k+1}))$$

Now b_{k+1} was a zero

$$\Rightarrow RHS = f(m(0)) = f(1) = 0$$
$$\Rightarrow LHS = RHS$$

So, in this case, we have p(k+1) is true.

Case 2 :

$$XOR(b_1, ..., b_k) = 0, b_{k+1} = 1$$
$$LHS = XOR(b_1, ..., b_{k+1}) = 0 \oplus 1 = 1$$
$$RHS = f\left(\prod_{i=1}^{k+1} m(b_i)\right) = f\left(\prod_{i=1}^{k} m(b_i).m(b_{k+1}\right)$$

$XOR(b_1, ..., b_k) = 0$ implies that $\prod_{i=1}^{k} m(b_i) = 1$ from the previous case

Substituting in the RHS, we get

$$RHS = f(1.m(b_{k+1})) = f(m(b_{k+1}))$$
$$= f(m(1) = f(-1) = 1$$
$$\Rightarrow LHS = RHS$$

Case 3 :

$$XOR(b_1, ..., b_k) = 1, b_{k+1} = 0$$
$$LHS = XOR(b_1, ..., b_{k+1}) = 1 \oplus 0 = 1$$
$$RHS = f\left(\prod_{i=1}^{k+1} m(b_i)\right) = f\left(\prod_{i=1}^{k} m(b_i).m(b_{k+1})\right)$$

XOR(b_1,...,b_{k+1}) = 1 implies that $\prod_{i=1}^{k} m(b_i) = -1$

Substituting in the RHS, we get

$$RHS = f((-1).m(b_{k+1})) = f(-m(0))$$
$$\Rightarrow f((-1) * (+1)) = f(-1) = 1$$
$$\Rightarrow LHS = RHS$$

Case 4 :

$$XOR(b_1, ..., b_k) = 1, b_{k+1} = 1$$
$$LHS = XOR(b_1, ..., b_{k+1}) = 1 \oplus 1 = 0$$
$$RHS = f\left(\prod_{i=1}^{k+1} m(b_i)\right) = f\left(\prod_{i=1}^{k} m(b_i).m(b_{k+1})\right)$$

Now, $XOR(b_1, ..., b_{k+1}) = 1$ implies that $\prod_{i=1}^{k}(b_i)$ = -1

Substituting in the RHS, we get

$$\begin{aligned} RHS &= f(-1.m(b_{k+1})) = f(-m(1)) \\ &= f((-1)*(-1)) = f(1) = 0 \\ &\therefore LHS = RHS \end{aligned}$$

So, p(k+1) is true

So, by the principle of mathematical induction, we have proved that,

$$XOR(b_1, ..., b_n) = f\left(\prod_{i=1}^{n} m(b_i)\right) \tag{9}$$

Question 2

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] ϵ $\mathbb{N}^9$. 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:

$$sign(\tilde{u}^T\tilde{x}).sign(\tilde{v}^T\tilde{x}).sign(\tilde{w}^T\tilde{x}) \tag{10}$$

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* ϵ $\mathbb{N}$, we always have

$$\prod_{i=1}^{n} sign(r_i) = sign\left(\prod_{i=1}^{n} r_i\right) \tag{11}$$

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 2

We have $r_1, r_2, ..., r_n$ as a set of real numbers where *n* ϵ $\mathbb{N}$. We assume a property P(t) such that

$$P(t) : \prod_{i=1}^{t} sign(r_i) = sign\left(\prod_{i=1}^{t} r_i\right) \tag{12}$$

Now lets check for P(1)

$$P(1) : LHS \Rightarrow \prod_{i=1}^{1} sign(r_i) = sign(r_1) \tag{13}$$

$$RHS \Rightarrow sign\left(\prod_{i=1}^{1}\right) = sign(r_1) \tag{14}$$

$$LHS = RHS = sign(r_1) \tag{15}$$

$\therefore$ p(1) is true.

Now lets check for some random natural number k<n, p(k) property. Assume that p(k) is true, we have

$$p(k) : \prod_{i=1}^{k} sign(r_i) = sign\left(\prod_{i=1}^{k} r_i\right) \tag{16}$$

Now lets check whether p(k+1) is true or not

$$p(k+1) : \prod_{i=1}^{k+1} sign(r_i) = sign\left(\prod_{i=1}^{k+1} r_i\right) \tag{17}$$

$$LHS = \prod_{i=1}^{k+1} sign(r_i) = sign(r_{k+1}).\left(\prod_{i=1}^{k} sign(r_i)\right) \tag{18}$$

$$RHS = sign\left(\prod_{i=1}^{k+1} r_i\right) = sign\left(r_{k+1}.\prod_{i=1}^{k} r_i\right) \tag{19}$$

let us assume two cases,

Case 1 : r_{k+1} is positive then,

$$sign\left(\prod_{i=1}^{k+1}\right) = +ve.sign\left(\prod_{i=1}^{k} r_i\right) \tag{20}$$

$$= sign(k_{k+1}).sign\left(\prod_{i=1}^{k} r_i\right) \tag{21}$$

Case 2 : r_{k+1} is negative then,

$$sign\left(\prod_{i=1}^{k+1} r_i\right) = -ve.sign\left(\prod_{i=1}^{k} r_i\right) \tag{22}$$

$$= sign(k_{k+1}).sign\left(\prod_{i=1}^{k} r_i\right) \tag{23}$$

$$So, RHS = sign\left(\prod_{i=1}^{k+1} r_i\right) = sign(r_{k+1}).sign\left(\prod_{i=1}^{k} r_i\right) \tag{24}$$

From our assumption of p(k), substituting in equation (24),

$$RHS = sign\left(\prod_{i=1}^{k+1} r_i\right) = sign(r_{k+1}).\left(\prod_{i=1}^{k} sign(r_i)\right) \tag{25}$$

So, p(k+1) is also coming true.

So, when p(k) is assumed to be true then p(k+1) is also coming true

So, by the principle of mathematical induction, p(n) is true for all n $\epsilon\mathbb{N}$

$$So, p(n) : \prod_{i=1}^{n} sign(r_i) = sign\left(\prod_{i=1}^{n} r_i\right) \tag{26}$$

is true for all n $\epsilon\mathbb{N}$

If any of $r_i\epsilon\mathbb{N}$ suppose $r_t = 0$ becomes zero, then as given in the question sign(0) = 0

$$LHS = \prod_{i=1}^{n} sign(r_i) = sign(r_1).sign(r_2)...sign(r_t)...sign(r_n) \tag{27}$$

$$Since,\ sign(r_t) = 0, LHS = 0 \tag{28}$$

Similarly,

$$RHS = sign\left(\prod_{i=1}^{n} r_i\right) = sign(r_1.r_2...r_t...r_n) \tag{29}$$

Now r_t = 0. So,

$$RHS = sign(r_1.r_2...0...r_n) = sign(0) = 0 \tag{30}$$

$$LHS = RHS \tag{31}$$

So, when one or more numbers is zero then also p(n) holds.

Question 3

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

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) \tag{32}$$

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 $\phi : \mathbb{R}^9 \rightarrow \mathbb{R}^D$ such that for any triple $(\tilde{u}, \tilde{v}, \tilde{w})$, there always exists a vector W$\epsilon\mathbb{R}^D$ such that for every $\tilde{x}\epsilon\mathbb{R}^9$, we have $(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = \mathrm{W}^T\phi(\tilde{x})$.

Solution 3

We know that for a single puff we can build a linear classifier w^{T}x+b or $\tilde{w}^T\tilde{x}$ here $\tilde{w}$ and $\tilde{x}$ are vector or dimension 9 (after hiding the bias term).
Now since we have three puffs we can try to model it in the form of one single linear classifier possibly in different dimension. (So that instead of training three different models we have to train a single linear model).
Now $\tilde{u}^T\tilde{x}$ is the scalar quantity whose sign tells the output of first puff, similarly $\tilde{v}^T\tilde{x}$ and $\tilde{w}^T\tilde{x}$ tells us the output of second and third puff.
In short we have to apply three different linear models.
Our goal is to make the product of the output of three feature vectors equal to the output of single linear model of possibly of different dimensionality so we don't need 3 different linear models instead now we have to work on a single linear model. $\mathrm{W}^T\phi(\tilde{x})$ is the linear model we have to create.

Now we know,

$$(\tilde{u}^T\tilde{x}) = \sum_{j=1}^{9} \tilde{u_j}\tilde{x_j} \tag{33}$$

$$= u_1x_1 + u_2x_2 + ... + u_9x_9 \tag{34}$$

$$(\tilde{v}^T\tilde{x}) = \sum_{j=1}^{9} \tilde{v_j}\tilde{x_j} \tag{35}$$

$$= v_1x_1 + v_2x_2 + ... + v_9x_9 \tag{36}$$

Now,

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}) = (u_1x_1 + u_2x_2 + ... + u_9x_9).(v_1x_1 + v_2x_2 + ... + v_9x_9) \tag{37}$$

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}) = \sum_{j=1}^{9}\sum_{k=1}^{9} \tilde{u_j}\tilde{v_k}\tilde{x_j}\tilde{x_k} \tag{38}$$

Similarly when we try to solve for 3 PUFs, if we expand the term

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = (\sum_{j=1}^{9} \tilde{u_j}\tilde{x_j}).(\sum_{j=1}^{9} \tilde{v_j}\tilde{x_j}).(\sum_{j=1}^{9} \tilde{w_j}\tilde{x_j}) \tag{39}$$

$$= \sum_{j=1}^{9}\sum_{k=1}^{9}\sum_{l=1}^{9} \tilde{u_j}\tilde{v_k}\tilde{w_l}\tilde{x_j}\tilde{x_k}\tilde{x_l} \tag{40}$$

$$So,\ (\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = u_1v_1w_1x_1x_1x_1+u_1v_1w_2x_1x_1x_2+u_1v_1w_3x_1x_1x_3+...+u_9v_9w_9x_9x_9x_9 \tag{41}$$

We could see that our new linear equation has 9*9*9 = 729 features where i,j,k $\epsilon[1,9]$. Now, we can see that $9^3 = 729$ dimensional function that maps $\tilde{x} = (\tilde{x_1}, \tilde{x_2}, ..., \tilde{x_9})$ to

$$\phi(\tilde{x}) = \begin{pmatrix} \tilde{x_1}\tilde{x_1}\tilde{x_1} & \tilde{x_1}\tilde{x_1}\tilde{x_9} & \tilde{x_1}\tilde{x_2}\tilde{x_9} & . & . & . & \tilde{x_9}\tilde{x_9}\tilde{x_9} \end{pmatrix} \tag{42}$$

$\phi(x)$ is of dimension 729*1 and now we can convert it into linear model of the form

$$(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = \mathrm{W}^T\phi(\tilde{x}) \tag{43}$$

$$where\ \mathrm{W} = \begin{pmatrix} \tilde{u_1}\tilde{v_1}\tilde{w_1} \\ \tilde{u_1}\tilde{v_1}\tilde{w_2} \\ . \\ . \\ . \\ \tilde{u_9}\tilde{v_9}\tilde{w_9} \end{pmatrix}$$

Dimentionality $(D) = (8+1)^3 = 729$, which depends on number of PUFs i.e. 3 and dimentionality of $\tilde{x} = 8 + 1 = 9$

and we have proved that there exists a way to map 9 dimensional vectors to D dimensional vectors as $\phi : \mathbb{R}^9 \rightarrow \mathbb{R}^D$ where D=729.

such that for any triple $(\tilde{u}, \tilde{v}, \tilde{w})$, there always exists a vector w$\epsilon\mathbb{R}^D$ i.e. w$\epsilon\mathbb{R}^{729}$ such that for every $\tilde{x}\epsilon\mathbb{R}^9$. we have $(\tilde{u}^T\tilde{x}).(\tilde{v}^T\tilde{x}).(\tilde{w}^T\tilde{x}) = \mathrm{W}^T\phi(\tilde{x})$.

Question 5

For the method you implemented, describe in your PDF report that 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 nad 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 5

To train our model we use learning rate (eta = 0.12) and lambda parameter (lambda_para = 0.0001) and initialize our weight metrics W with zeros as this gives the best accuracy to our model. Further with each iteration our learning rate will decrease by $\sqrt{t}$ where t is the iteration number. So that our model will converge easily.

We learned these hyperparameters by randomly using different values for each parameter it is more like random search huyperparameter tuning.

We assign different values for eta and lambda_para in search space with random values and predict using the model for each setting of hyperparameter. Our metric to find our hyperparameter. So we find on what setting of hyperparameter our model performs best and give high accuracy or produces minimum classification error.

eta	**lambda_para**	**Validation Accuracy**
0.001	0.5	0.7595
0.12	0.0001	0.9895
0.001	0.2	0.822

An instance from our parameter turning as shown above tells us that at values eta=0.12 and lambda_para=0.0001 gives best validation accuracy of 0.9895

Question 6

Plot the convergence curves in your PDF report offered by your chosen method as we do in lecture notebooks. This 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 6