CS771: Introduction to Machine Learning
Assignment 1

Group Name : Perceptron

Problem 1 :

Given, set of binary digits b_1, b_2, b_3,......., b_n such that $n \in N$ where, each $b_i \in \{0, 1\}$ for i =1,2......,n
Now, we know that for any set of binary digits b_1, b_2,....b_n

$$XOR(b_1, b_2, ...b_n) = \begin{cases} 1, & \text{if there are odd number of } b_i \text{ having value 1} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Now, if we choose a function m such that

$$m(0) = 1 \text{ and } m(1) = -1 \text{ or } m(x) = 1 - 2x \tag{2}$$

$$f(-1) = 1 \text{ and } f(1) = 0 \text{ or } f(x) = (1 - x)/2 \tag{3}$$

Then,

$$m(b_i) = \begin{cases} -1, & \text{if } b_i = 1 \\ 1, & \text{else if } b_i = 0 \end{cases} \tag{4}$$

This means that

$$\prod m(b_i) = \begin{cases} -1, & \text{if there are odd number of } b_i \text{ having value 1} \\ 1, & \text{if there are even number of } b_i \text{ having value 1} \end{cases} \tag{5}$$

Applying function f given in equation (3) on equation (5) we get,

$$f\left(\prod m(b_i)\right) = \begin{cases} 1 & \text{if there are odd number of } b_i \text{ having value 1} \\ 0 & \text{if there are even number of } b_i \text{having value 1} \end{cases} \tag{6}$$

From equation (1) and (6) we can conclude that

$$XOR(b_1, b_2, ..., b_n) = f\left(\prod m(b_i)\right) \tag{7}$$

for the function f and m as defined in equation 2 and equation 3

Problem 2 :

Given, we are given set of n real numbers r_1, r_2,.....,r_n such that $sign(0) = 0$
To Prove:

$$\prod_{i=1}^{n} sign(r_i) = sign\left(\prod_{i=1}^{n} r_i\right) \text{ for any } n \in N \tag{8}$$

Case 1: when value of atleast one of the r_i is zero. let us say $r_k = 0$ where $1 < k < n$ In this case, L.H.S is given by,

$$\left(\prod_{i=1}^{k-1} sign(r_i)\right) \cdot sign(r_k) \cdot \left(\prod_{i=k+1}^{n} sign(r_i)\right) \tag{9}$$

Now, since $r_k = 0$ As product of real number with zero is 0.Therefore, $sign(r_k)$ = 0 (given)
Therefore, L.H.S = 0

R.H.S. is given by

$$sign\left(\prod_{i=1}^{n} r_i\right) = sign\left(\left(\prod_{i=1}^{k-1} r_i\right) \cdot (r_k) \cdot \left(\prod_{i=k+1}^{n} r_i\right)\right) \tag{10}$$

Now, since $r_k = 0$ Therefore,

$$\prod_{i=1}^{n} r_i = 0 \tag{11}$$

Since $sign(0) = 0$ (Given). Therefore,

$$sign\left(\prod_{i=1}^{n} r_i\right) = 0 \tag{12}$$

From equation(9) and equation (12) we have L.H.S = R.H.S. Proved
Case 2: $r_i \neq 0$ for i =1,2.......,n for any $n \in N$
R.H.S is given by,

$$sign\left(\prod_{i=1}^{n} r_i\right) = sign\left(\prod_{i=1}^{n} |r_i| . \prod_{i=1}^{n} sign(r_i)\right) \tag{13}$$

Now, $|r_i|$ is always positive. Therefore,

$$\prod_{i=1}^{n} |r_i| \text{ is positive} \tag{14}$$

Therefore, equation (13) can be rewritten as

$$sign\left(\prod_{i=1}^{n} sign(r_i)\right) \tag{15}$$

And Thus $R.H.S$ is same as

$$sign\,(L.H.S.) \tag{16}$$

Now, $L.H.S$ is given by,

$$\prod_{i=1}^{n} sign(r_i) \tag{17}$$

We can see from above equation (17) that $L.H.S$ can be either 1 or -1
Thus, if $L.H.S$ = 1 then, $R.H.S = sign(L.H.S)$ = 1
Else if $L.H.S$ =-1 then, $R.H.S$ =$sign(L.H.S)$ =-1
Thus, in Case 2 also $L.H.S$ =$R.H.S$
From the above two cases we can conclude that $L.H.S$ =$R.H.S$ [proved]

Problem 3 :

Given, expression

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right)\cdot\left(\mathbf{w}^T\mathbf{x}\right) \tag{18}$$

We need to find a $\mathbf{W}$ such that $\mathbf{W} \in R^D$ and a function ϕ: $\mathcal{R}^9 \rightarrow \mathcal{R}^D$ such that for every vector $\mathbf{x} \in R^9$, we have,

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right)\cdot\left(\mathbf{w}^T\mathbf{x}\right) = \mathbf{W}^T\phi(\mathbf{x}) \tag{19}$$

Let us first solve for case when we have two terms in the $L.H.S$ i.e.

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right) \tag{20}$$

Now, the above equation can be rewritten as

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right) = \left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(\sum_{j=1}^{9} v_j x_j\right) \tag{21}$$

Now,

$$\left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(\sum_{j=1}^{9} v_j x_j\right) = \left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(v_1x_1 + v_2x_2 + ... + v_9x_9\right) \tag{22}$$

$$= \left(u_1x_1 + u_2x_2 + .. + u_9x_9\right)v_1x_1 + \left(u_1x_1 + u_2x_2 + .. + u_9x_9\right)v_2x_2 + ... \tag{23}$$

$$= u_1x_1v_1x_1 + u_1x_1v_2x_2 + ... + u_2x_2v_1x_1 + u_2x_2v_2x_2 + ... + u_9x_9v_1x_1 + ... + u_9x_9v_9x_9 \tag{24}$$

Thus the above equation can be rewritten as,

$$= \sum_{i=1}^{9}\sum_{j=1}^{9} u_i x_i v_j x_j \tag{25}$$

$$= \sum_{i=1}^{9}\sum_{j=1}^{9} u_i v_j x_i x_j \tag{26}$$

From equation (22) and equation(26), we get

$$\left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(\sum_{j=1}^{9} v_j x_j\right) = \sum_{i=1}^{9}\sum_{j=1}^{9} u_i v_j x_i x_j \tag{27}$$

Thus, from above equation we see that we can create 81 dimensional function that maps $\mathbf{x} = (x_1, ..., x_9)$ to $\phi(\mathbf{x}) = (x_1x_1, x_1x_2.., x_1x_9, x_2x_1, ..., x_9x_9)$ Thus, from equation (21) and equation (26) we get

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right) = \mathbf{W}^T\phi(\mathbf{x}) \tag{28}$$

Where,

$$\mathbf{W} = (u_1v_1, u_1v_2, .., u_1v_9, u_2v_1, ..., u_9v_9) \tag{29}$$

Now, Working on similar lines as above we can proceed and solve expression (18)

$$\left(\mathbf{u}^T\mathbf{x}\right)\cdot\left(\mathbf{v}^T\mathbf{x}\right)\cdot\left(\mathbf{w}^T\mathbf{x}\right) = \left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(\sum_{j=1}^{9} v_j x_j\right)\cdot\left(\sum_{k=1}^{9} w_k x_k\right) \tag{30}$$

Now, using the result derived in equation(27), we can rewrite the equation as

$$= \left(\sum_{i=1}^{9} u_i x_i\right)\cdot\left(\sum_{j=1}^{9}\sum_{k=1}^{9} v_j w_k x_j x_k\right) \tag{31}$$

Again, applying the result derived in equation(27), we can rewrite the above equation as

$$= \left(\sum_{i=1}^{9} \sum_{j=1}^{9} \sum_{k=1}^{9} u_i v_j w_k x_i x_j x_k \right) \tag{32}$$

Thus, from above equation we see that we can create 729 dimensional function that maps $\mathbf{x} = (x_1, ..., x_9)$ to $\phi(\mathbf{x}) = (x_1x_1x_1, x_1x_1x_2.., x_1x_1x_9, x_1x_2x_1, ..., x_9x_9x_9)$ Thus, from equation (30) and equation (32) we get

$$\left(\mathbf{u}^T\mathbf{x}\right) \cdot \left(\mathbf{v}^T\mathbf{x}\right) \cdot \left(\mathbf{w}^T\mathbf{x}\right) = \mathbf{W}^T\phi(\mathbf{x}) \tag{33}$$

Where,

$$\mathbf{W} = (u_1v_1w_1, u_1v_1w_2, .., u_1v_1w_9, u_1v_2w_1, ..., u_9v_9w_9) \tag{34}$$

Thus, vector $\mathbf{W} \in R^{729}$ as given in equation(34) and a function ϕ: $\mathcal{R}^9 \rightarrow \mathcal{R}^{729}$ as given above exists such that

$$\left(\mathbf{u}^T\mathbf{x}\right) \cdot \left(\mathbf{v}^T\mathbf{x}\right) \cdot \left(\mathbf{w}^T\mathbf{x}\right) = \mathbf{W}^T\phi(\mathbf{x}) \tag{35}$$

Problem : 5

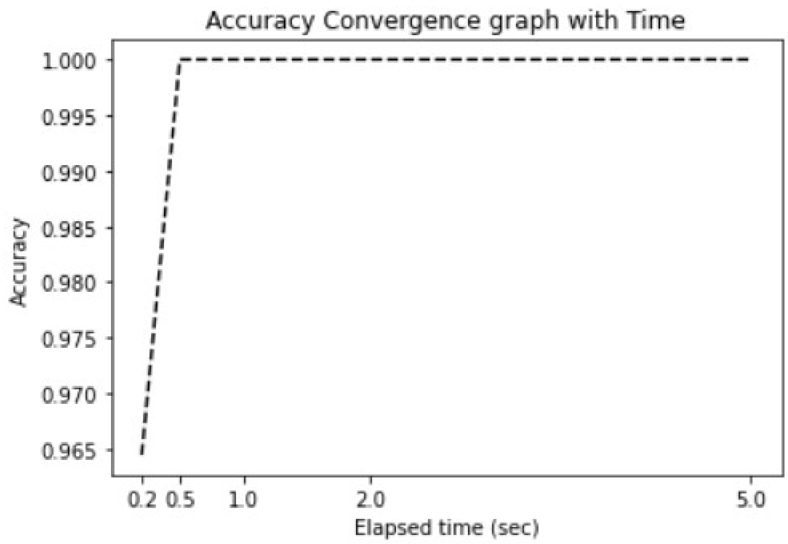
Method Used For Optimization : **Stochastic Gradient Descent**

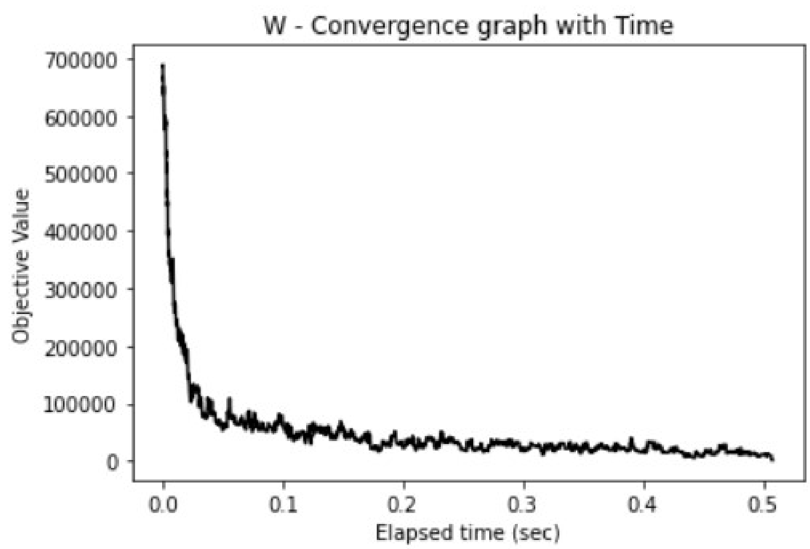
Policy used to choose the next data point from the Dataset : **Sequential**

<u>Hyperparameters</u> :

- Learning Rate or Step Length : **0.027**
  - We tried **constant** step length of 0.1, 0.01, 0.02, 0.027, turned out 0.027 was giving the best results among all. Moreover we tried making it linear with initial learning rate - 1, 2, 2.5, 3 as well but constant turned out to be the best.
- C : **3**
  - We tried C to be 1, 2, 4, 3, 2.5, 2.9, 3.1 of which 3 was giving the best results among all.
- Regularizer (r) : $\mathbf{10^{-5}}$
  - We tried regularizer (r) to be 10^{-1}, 10^{-3}, 10^{-4}, 10^{-5} of which 10^{-5} was giving the best results among all.

Problem : 6


Accuracy Convergence graph with Time
Accuracy
1.000
0.995
0.990
0.985
0.980
0.975
0.970
0.965
0.2
0.5
1.0
2.0
5.0
Elapsed time (sec)


W - Convergence graph with Time
Objective Value
700000
600000
500000
400000
300000
200000
100000
0
0.0
0.1
0.2
0.3
0.4
0.5
Elapsed time (sec)