Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2022-23-1

QUESTION 1

Assignment Number: 1
Group Name: is_this_fft
Group Number: 29
Group Members: Abhishek Pardhi, Ankur Kumar, Parth Maniar, Sahil Bansal, Suket Raj
Date: September 17, 2022

Problem 1: XOR is a product

We need to show that there exists a mapping m from binary {0,1} to signs {-1,1} with map m and then another with map $f : \{-1, 1\} \rightarrow \{0, 1\}$ such that for a set of binary digits $b_1, b_2...b_n$ for some $n \epsilon N$, we have

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

Consider the mapping m:

$$m(b) = \begin{cases} -1, & \text{if } b = 1 \\ +1, & \text{if } b = 0 \end{cases} \tag{1}$$

and the mapping f:

$$f(x) = \begin{cases} 0, & \text{if } x = +1 \\ 1, & \text{if } x = -1 \end{cases} \tag{2}$$

The function XOR returns 1 if we provide odd number of 1s in the binary sequence and shall ignore the number of 0s in the same. We exploit the similar relationship between product of 1s and -1s where the number of 1s dont determine the answer but the number of -1s do.
Let us consider a general case where in a binary sequence(S) we have x number of 0s and y number of 1s, We say that,

$$XOR(S) = \begin{cases} 0, & \text{if } y \text{ is even} \\ 1, & \text{if } y \text{ is odd} \end{cases} \tag{3}$$

Suppose we map $0 \rightarrow 1$ and $1 \rightarrow -1$, the product of x 1s and y -1s shall be,

$$prod(S) = \begin{cases} -1, & \text{if } y \text{ is even} \\ 1, & \text{if } y \text{ is odd} \end{cases} \tag{4}$$

which turns out very similar, so now if we use mapping from (1) to encode the entries and multiply all them and later decode the answer using mapping (2), we get the same output as the XOR function. Hence, we can represent XOR of a binary sequence as a product of its mapped entries.

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2022-23-1

QUESTION

2

Assignment Number: 1
Group Name: is_this_fft
Group Number: 29
Group Members: Abhishek Pardhi, Ankur Kumar, Parth Maniar, Sahil Bansal, Suket Raj
Date: September 17, 2022

Problem 2: Product of sign is sign of product

To prove:

$$\prod_{i=1}^{n} sign(b_i) = sign\left(\prod_{i=1}^{n} b_i\right), \forall\, b_i \in \mathbb{R}$$

Le $n = 2$ be the base case-

$$sign(ab) = sign(a)sign(b)$$

Case-1: $a > 0$ & $b > 0 \implies sign(ab) = 1 = 1 \cdot 1 = sign(a)sign(b)$
Case-2: $a < 0$ & $b < 0 \implies sign(ab) = 1 = -1 \cdot -1 = sign(a)sign(b)$
Case-3: $a < 0$ & $b > 0 \implies sign(ab) = -1 = -1 \cdot 1 = sign(a)sign(b)$
Case-4: $a > 0$ & $b < 0 \implies sign(ab) = 1 = 1 \cdot -1 = sign(a)sign(b)$
Case-5: at least one of a, b is $0 \implies ab = 0 \implies sign(ab) = 0 = sign(a)sign(b)$
Let the claim be true $\forall\, n \leq k$ then by induction hypothesis-

$$\prod_{i=1}^{k+1} sign(b_i) = \left(\prod_{i=1}^{k} sign(b_i)\right) sign(b_{k+1}) = sign(\prod_{i=1}^{k} b_i) sign(b_{k+1})$$

now as proved for $n = 2$,

$$= sign(\prod_{i=1}^{k} b_i \cdot b_{k+1}) = sign(\prod_{i=1}^{k+1} b_i)$$

Hence by induction claim is true.

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2022-23-1

QUESTION

3

Assignment Number: 1
Group Name: is_this_fft
Group Number: 29
Group Members: Abhishek Pardhi, Ankur Kumar, Parth Maniar, Sahil Bansal, Suket Raj
Date: September 17, 2022

Problem 3: Calculation of Dimensionality

We need to show that the quantity

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

can be expressed as a linear model but in a higher dimensional space. Also the dimensions of the model is dependent upon only the number of PUFs and the dimensionality of the vector $\tilde{x}$.

Let the dimensionality of $\tilde{x}$ be D and number of PUFs be n. Now lets assume that we can express n such linear models of D dimensionality into a new linear model of dimensionality D^n. The case where n=1 is trivial as the origional model is the required model with dimensionality D^1. Suppose a model with k PUFs can be expressed as a new model, i.e.

$$\prod_{i=1}^{k}(V_i^\top \tilde{x}) = W_k^\top \phi_k(\tilde{x})$$

where V_i is the model vector for each linear model and $W_k \, \epsilon \, \mathbb{R}^{D^k}$ and $\phi_k : \mathbb{R}^D \rightarrow \mathbb{R}^{D^n}$.
Let us look at the LHS closely, it should look like this:

$$(\sum_{a=1}^{D} V_{1_a}^\top \tilde{x_a})(\sum_{b=1}^{D} V_{2_b}^\top \tilde{x_b})..total\ k\ times...(\sum_{c=1}^{D} V_{k_c}^\top \tilde{x_c})$$

$$\sum_{a=1}^{D}\sum_{b=1}^{D}..total\ k\ times..\sum_{c=1}^{D}(V_{1_a}V_{2_b}..V_{k_c}\tilde{x_a}\tilde{x_b}..\tilde{x_c})$$

this implies that the ϕ_k maps

$\tilde{x} = \{x_1, x_2, ...x_n\}\ to\ \phi_k(\tilde{x}) = \{\tilde{x_1}..\tilde{x_1}\tilde{x_1}, \tilde{x_1}..\tilde{x_1}\tilde{x_2}, ..\tilde{x_1}..\tilde{x_1}\tilde{x_n}, \tilde{x_1}..\tilde{x_2}\tilde{x_1}, ..., \tilde{x_n}..\tilde{x_n}\tilde{x_n}\}$

where each term is of size k and similarly we can comment on the model vector W_k.
Now if we introduce another linear model into the picture V_{k+1} and multiply it to the LHS we get the following

$$(\sum_{t=1}^{D} V_{k+1_t}^\top \tilde{x_t})\sum_{a=1}^{D}\sum_{b=1}^{D}..total\ k\ times..\sum_{c=1}^{D}(V_{1_a}V_{2_b}..V_{k_c}\tilde{x_a}\tilde{x_b}..\tilde{x_c})$$

$$\sum_{a=1}^{D}\sum_{b=1}^{D}..total\ k+1\ times..\sum_{t=1}^{D}(V_{1_a}V_{2_b}..V_{k+1_t}\tilde{x_a}\tilde{x_b}..\tilde{x_t})$$

and we define a new mapping here $\phi_{k+1} : \mathbb{R}^D \rightarrow \mathbb{R}^{D^{k+1}}$ which maps

$\tilde{x} = \{x_1, x_2, ...x_n\}\ to\ \phi_k(\tilde{x}) = \{\tilde{x_1}..\tilde{x_1}\tilde{x_1}, \tilde{x_1}..\tilde{x_1}\tilde{x_2}, ..\tilde{x_1}..\tilde{x_1}\tilde{x_n}, \tilde{x_1}..\tilde{x_2}\tilde{x_1}, ..., \tilde{x_n}..\tilde{x_n}\tilde{x_n}\}$

where each term of the mapping is of size k+1, and a similar procedure goes for the construction for the new model vector W_{k+1}, this implies that

$$W_k^\top \phi_k(\tilde{x}).(V_{k+1}\tilde{x}) = W_{k+1}^\top \phi_{k+1}(\tilde{x})$$

Hence by Induction we proved that for any n the following shall hold

$$\prod_{i=1}^{n}(V_i^\top \tilde{x})^\top \phi(\tilde{x}) = W^\top \phi(\tilde{x})$$

where V_i is the model vector for each linear model and W $\epsilon \, \mathbb{R}^{D^n}\ and\ \phi : \mathbb{R}^D \rightarrow \mathbb{R}^{D^n}$.
In our case D=9 and n=3, hence we can represent the 3 linear models in a single model of 729 dimensions.

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2022-23-1

Assignment Number: 1
Group Name: is_this_fft
Group Number: 29
Group Members: Abhishek Pardhi, Ankur Kumar, Parth Maniar, Sahil Bansal, Suket Raj
Date: September 17, 2022

QUESTION

5

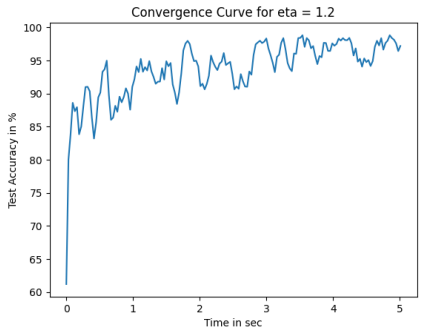
Problem 5: Model Report

0.1 Optimization Method

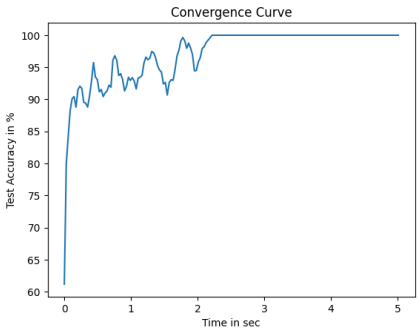
We used *vanilla* Gradient Descent.
Gradient Descent computes gradients of the loss function, and update parameters in the opposite direction of gradient to decrease loss.
Vanilla GD is the simplest version of GD and updates all coordinates on complete dataset at once.
We experimented with Stochastic Dual Coordinate Maximization (SDCM) but it was giving very low accuracies in the range of 80-85%, so we switched to GD

0.2 Step length

The step length used is $\eta = \frac{c}{t}$ where c is some constant and t is time. Upon testing our model for multiple values of c from 0.5 to 2, we saw that the accuracy was good enough between any value in $[0.5, 1]$ but decreased gradually as c increased from 1. So we took $c = 0.8$ as our final parameter for the step length. The below graph shows that the model doesn't converge in the given amount of time for $\eta = \frac{1.2}{t}$ but its able to converge for $\eta = \frac{0.8}{t}$.



(a) $\eta = \frac{1.2}{t}$



(b) $\eta = \frac{0.8}{t}$

0.3 Penalty parameter C of CSVM

For small values of C, we saw under-fitting so we chose a value that wasn't much small, i.e., $C = 5.0$.

Assignment Number: 1
Group Name: is_this_fft
Group Number: 29
Group Members: Abhishek Pardhi, Ankur Kumar, Parth Maniar, Sahil Bansal, Suket Raj
Date: September 17, 2022

Problem 6: Plot of convergence curves

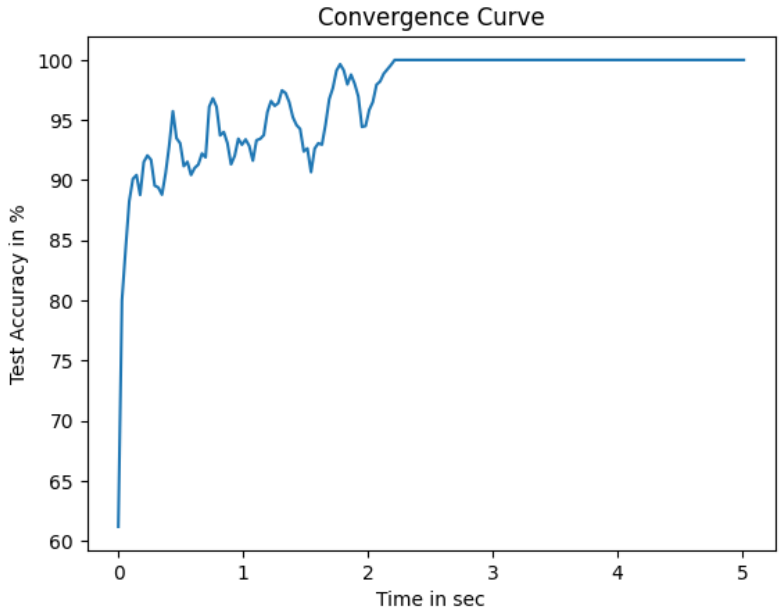


Figure 2: Test classification accuracy with time for $\eta = 0.8$

The plot is converging to 100% after $t = 2.2sec$.