CS771 : Assignment 1

1. Rahul Rustagi (200756)
Department of Aerospace Engineering
rrustagi20@iitk.ac.in

2. Arnav Shendurnikar (200929)
Department of Chemical Engineering
sarnav20@iitk.ac.in

3. Rachit Khamsera (200747)
Department of Economics Engineering
rachit20@iitk.ac.in

4. Arihant Jha (200181)
Department of Aerospace Engineering
arihantj20@iitk.ac.in

5. P Madhav (200657)
Department of Chemical Engineering
pmadhav20@iitk.ac.in

6. Siddesh Bharat Hazare (200976)
Department of Aeropsace Engineering
siddeshbh20@iitk.ac.in

Abstract

The file contains the submission of Assignment 1 of CS771 from the team, Melbo's Guys. The answers to the first, second and fourth question have been answered in this pdf file. The answer to the third question is answered in the "submit.py" file.

Team Name : **Melbo's Guys**

INDEX

Preprint. Under review.

Solution 1

We discuss the proof and derive the mathematical expression of the final output y.

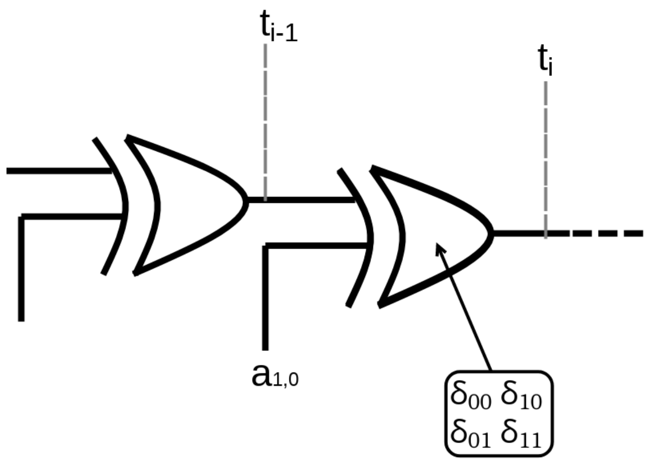

Figure 1: XORRO Model

Objective of our model is to classify which of the two XORROs has higher frequency.
The frequency of a XORRO is given by

$$f_{XORRO} = \frac{1}{t_l + t_u}$$

Frequency difference between upper and lower XORROs i.e. $\Delta_i = f_i^u - f_i^l$ decides the output value.
$\Delta_i > 0 \implies$ counter value 1
$\Delta_i < 0 \implies$ counter value 0
Since working with $\frac{1}{t}$ introduces non-linearity in function relation of classifier, we choose to work with $\Delta_i = t_i^l - t_i^u$
To formulate the mathematical expression of the linear classifier refer the above diagram.

- Consider an i_{th} XOR in the XORRO Model. Let t_{i-1} be the total delay accumulated or time elapsed from the input to first XOR in the model to the signal reaching the i_{th} XOR.
- t_i is, mathematically, t_{i-1} + delay acquired or time elapsed in the i_{th} XOR.
- As we know, if a_i is high, then XOR acts as an inverter and if its low, then it acts like an OR gate.
- We can write the following expression,

$$t_i = t_{i-1} + a_i(\delta_{11}^i + \delta_{01}^i) + (1 - a_i)(\delta_{10}^i + \delta_{00}^i) \tag{1}$$

- It is important to note that all values and permutations of a_i and δ_{jk}^i have to be considered to get the total possible delay in a XORRO model. This observation greatly simplifies the derivation of linear relation between the input and output of the XORRO model.
- Simplifying equation (1),

$$t_i = t_{i-1} + a_i(\delta_{11}^i + \delta_{01}^i - \delta_{10}^i - \delta_{00}^i) + (\delta_{10}^i + \delta_{00}^i) \tag{2}$$

- Define Δ as the difference of time between upper and lower XORRO model. Hence,

$$\Delta_i = t_i^l - t_i^u \tag{3}$$

- The time difference can now be written as

$$\Delta_i = \Delta_{i-1} + a_i[(\delta_{11}^i + \delta_{01}^i - \delta_{10}^i - \delta_{00}^i)^l - (\delta_{11}^i + \delta_{01}^i - \delta_{10}^i - \delta_{00}^i)^u] \\ +[(\delta_{10}^i + \delta_{00}^i)^l - (\delta_{10}^i + \delta_{00}^i)^u] \tag{4}$$

- Let

$$\alpha_i = [(\delta_{11}^i + \delta_{01}^i - \delta_{10}^i - \delta_{00}^i)^l - (\delta_{11}^i + \delta_{01}^i - \delta_{10}^i - \delta_{00}^i)^u] \tag{5}$$

$$\beta_i = [(\delta_{10}^i + \delta_{00}^i)^l - (\delta_{10}^i + \delta_{00}^i)^u] \tag{6}$$

- Therefore the final equation becomes:

$$\Delta_i = \Delta_{i-1} + a_i\alpha_i + \beta_i \tag{7}$$

- Solving the recurrence,

$\Delta_0 = 0$ (signifying initial delay in input to first XOR)
$\Delta_1 = a_1\alpha_1 + \beta_1$
$\Delta_2 = a_1\alpha_1 + \beta_1 + a_2\alpha_2 + \beta_2$
$\Delta_3 = a_1\alpha_1 + \beta_1 + a_2\alpha_2 + \beta_2 + a_3\alpha_3 + \beta_3$
.
.
.
$\Delta_R = w^T x + b_R$
More Specifically,

$$w_i = \alpha_i \quad x_i = a_i \quad b_i = \sum_{i=1}^{R} \beta_i$$

$$w = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_R \end{bmatrix} x = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_R \end{bmatrix} b = \begin{bmatrix} \beta_1 \\ \beta_1 + \beta_2 \\ \vdots \\ \sum_{i=1}^{R} \beta_i \end{bmatrix} \tag{8}$$

x is the challenges that equal to the input bits

w is the weights of the linear classifier

Hence $\Delta_{63} = w^T x + b_{63}$
If $\Delta_{63} < 0$ response = 0
If $\Delta_{63} > 0$ response = 1
Solution will be

$$\frac{1 + sig(w^T x + b)}{2}$$

Therefore,

$$\phi(c) = c$$

Solution 2

Given

$$
\begin{aligned}
a &= [a_0, a_1, ..a_{R-1}] \\
p &= [p_0, p_1, ..p_{S-1}] \\
q &= [q_0, q_1, ..q_{S-1}]
\end{aligned}
$$

Number of Xorro's = 2^S

Therefore number of possible combinations = number of ways to pick the first Xorro * number of ways to pick the second Xorro

Let it be M

$$M = 2^S(2^S - 1) \tag{9}$$

from q.1. we have

$$\Delta_{63} = w^T x + b \tag{10}$$

The above model will predict the result for any given 2 XORROs

So let p & q be the XORROs selected from the set of 2^S XORROs

Therefore

$$\Delta^{(p,q)} = w_{p,q}^T x + b_{p,q} \tag{11}$$

where

$$w_{p,q}^T = [\alpha_1^{p,q}, \alpha_2^{p,q}, ...\alpha_{64}^{p,q}] \tag{12}$$

$$b_{p,q}^T = [\beta_1^{p,q}, \beta_1^{p,q} + \beta_2^{p,q}, ... \sum_{i=1}^{R} \beta_i^{p,q}]$$

(13)

where

$$\alpha_i^{p,q} = [(\delta_{11}^{(i)q} + \delta_{01}^{(i)q} - \delta_{10}^{(i)q} - \delta_{00}^{(i)q}) - (\delta_{11}^{(i)p} + \delta_{01}^{(i)p} - \delta_{10}^{(i)p} - \delta_{00}^{(i)p})] \tag{14}$$

$$\beta_i^{p,q} = (\delta_{10}^{(i)q} + \delta_{00}^{(i)q}) - (\delta_{10}^{(i)p} + \delta_{00}^{(i)p}) \tag{15}$$

p^{th} Xorro is selected by mux 0

q^{th} Xorro is selected by mux 1

Therefore for $M = 2^S(2^S - 1)$ combinations we will have M models

Therefore $\forall$ (p,q) $\in$[0, $2^S - 1$]

Response will be given by

$$\frac{1 + sig(w_{p,q}^T x + b_{p,q})}{2} \tag{16}$$

We also note (p,q) pair will have opposite response to (q,p) i.e.,

$$\Delta^{p,q} = -\Delta^{q,p} \tag{17}$$

Proving that $\Delta^{p,q} = -\Delta^{q,p}$

Derivation :

$$\Delta^{p,q} = w_{p,q}^T x + b_{p,q}$$
$$w_{p,q}^T = [\alpha_0^{p,q}, \alpha_1^{p,q}, \dots, \alpha_{63}^{p,q}]$$
$$b^T = [\beta_0^{p,q}, \beta_0^{p,q} + \beta_1^{p,q}, \dots, \sum_{i=1}^{R} \beta_i^{p,q}]$$
$$\alpha_i^{p,q} = [(\delta_{01}^{(i)q} + \delta_{11}^{(i)q} - \delta_{10}^{(i)q} - \delta_{00}^{(i)q}) - (\delta_{01}^{(i)p} + \delta_{11}^{(i)p} - \delta_{10}^{(i)p} - \delta_{00}^{(i)p})]$$

Now,

$$\alpha_i^{q,p} = [(\delta_{01}^{(i)p} + \delta_{11}^{(i)p} - \delta_{10}^{(i)p} - \delta_{00}^{(i)p}) - (\delta_{01}^{(i)q} + \delta_{11}^{(i)q} - \delta_{10}^{(i)q} - \delta_{00}^{(i)q})]$$

Hence,

$$\alpha_i^{p,q} = -\alpha_i^{q,p}$$

Similarly, we can obtain $\beta_i^{p,q}$, $\beta_i^{q,p}$:

$$\beta_i^{p,q} = (\delta_{10}^{(i)q} + \delta_{00}^{(i)q - \delta_{10}^{(i)p} - \delta_{00}^{(i)p}})$$
$$\beta_i^{q,p} = (\delta_{10}^{(i)p} + \delta_{00}^{(i)p - \delta_{10}^{(i)q} - \delta_{00}^{(i)q}})$$
$$\beta_i^{p,q} = -\beta_i^{q,p}$$

And for $w_{p,q}$, $w_{q,p}$:

$$\begin{aligned} w_{p,q}^T &= [\alpha_0^{p,q}, \alpha_1^{p,q}, \dots, \alpha_{63}^{p,q}] \\ &= [-\alpha_0^{q,p}, -\alpha_1^{q,p}, \dots, -\alpha_{63}^{q,p}] \\ &= (-1)[\alpha_0^{p,q}, \alpha_1^{p,q}, \dots, \alpha_{63}^{p,q}] \\ &= -w_{q,p}^T \end{aligned}$$

Similarly,

$$b_{p,q} = -b_{q,p}$$

Therefore, we can say:

$$\begin{aligned} \Delta^{p,q} &= w_{p,q}^T x + b_{p,q} \\ &= -w_{q,p}^T x - b_{q,p} \\ &= (-1)(w_{q,p}^T x + b_{q,p}) \\ \Delta^{p,q} &= -\Delta^{q,p} \end{aligned}$$

Therefore models can be further modified:

If $p < q$

$$\frac{1 + sig(w_{p,q}^T x + b_{p,q})}{2} \tag{18}$$

If $p > q$

$$\frac{1 + sig(-(w_{p,q}^T x + b_{p,q}))}{2} \tag{19}$$

Therefore we will use half the models used previously i.e., M = $\frac{2^S(2^S-1)}{2} = 2^{S-1}(2^S-1)$

Solution 4

Linear SVC			
Hyperparameters	Type	Training time	Test accuracy
Loss Function	Hinge	3.5407543ll3	93.954
	Square Hinge	3.75315624	**94.741**
C Hyperparameter	High (1e6)	3.2896592	93.8504
	Low (1e-6)	2.2915173	76.1525
	Medium (1.0)	3.75315624	**94.741**
Tol Hyperparameter	High (1e6)	2.28088148	87.3875
	Low (1e-6)	3.75315624	**94.741**
	Medium (1.0)	2.53287253	94.58949

Logistic Regression			
Hyperparameters	Type	Training time	Test accuracy
C Hyperparameter	High (1e6)	5.1021398	94.5575
	Low (1e-6)	3.098767	76.1875
	Medium (1.0)	4.42149	93.9175
Tol Hyperparameter	High (1e6)	2.4948314	53.7
	Low (1e-6)	4.42149	93.9175
	Medium (1.0)	3.397926	93.78
Penalty	L1	2.676095	93.5389
	L2	4.42149	93.9175

Therefore from all the experiments we can infer that training time and test accuracy is comparatively lesser for Linear SVC than Logistic Regression. Also by lowering the C Hyperparameter, training time decreases but test accuracy also decreases and vice versa. So an optimum value is chosen. Test accuracy increases with a decrease in Tol Hyperparameter, and it also increases training time.