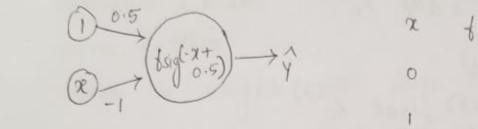
Name: Tharani Maareeivaannan

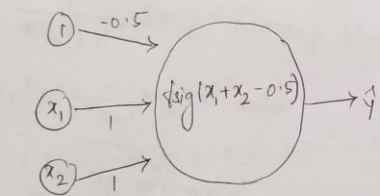
ID: 2575198

1. NOT gate



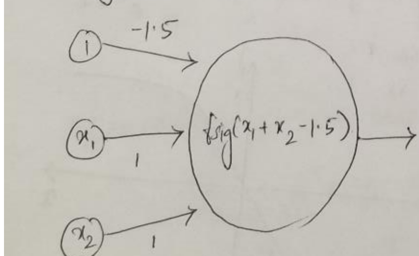
x	$f_{sig}(-x+0.5)$	$\hat{y}$ O/P
0	$f_{sig}(-0+0.5) = 0.5$	1
1	$f_{sig}(-1+0.5) = -0.5$	0

OR gate



x_1	x_2	$f_{sig}(x_1+x_2-0.5)$	$\hat{y}$ O/P
0	0	$f_{sig}(0+0-0.5) = f_{sig}(-0.5)$	0
0	1	$f_{sig}(+0.5)$	1
1	0	$f_{sig}(+0.5)$	1
1	1	$f_{sig}(+1.5)$	1

AND gate



x_1	x_2	$f_{sig}(x_1+x_2-1.5)$	$\hat{y}$
0	0	$f_{sig}(0+0-1.5) = f_{sig}(-1.5)$	0
0	1	$f_{sig}(-0.5)$	0
1	0	$f_{sig}(-0.5)$	0
1	1	$f_{sig}(+0.5)$	1

2. Vector equation in the forward propagation

$$z^{(1)} = X$$

$$T^{(2)} = \beta^{(1)T} z^{(1)}$$

$$z^{(2)} = f_{sig} \cdot T^{(2)} \text{ \& add } z_0^{(2)}$$

$$T^{(3)} = \beta^{(2)T} z^{(2)}$$

$$z^{(3)} = f_{sig} \cdot T^{(3)} \text{ \& add } z_0^{(3)}$$

$$T^{(4)} = \beta^{(3)T} z^{(3)}$$

$$z^{(4)} = \hat{Y} = f_{sig} \cdot T^{(4)}$$

3. (a) characteristic curve for sigmoid function $f(x) = \frac{1}{1+e^{-x}}$

when $x=0 \Rightarrow \frac{1}{1+e^{-0}} = 1$

$x = 4 = \frac{1}{1+e^{-4}} = 1$

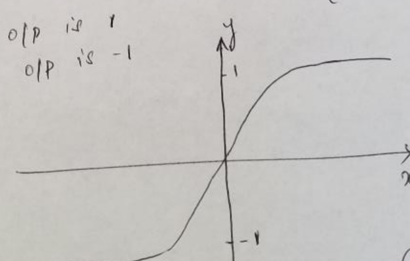
when i/p +ve o/p is 1

when -ve o/p is 0

characteristic curve for hyperbolic tangent function $f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

when i/p is +ve o/p is 1

when i/p is -ve o/p is -1



(b) When the ~~Starting~~ starting values of weight are very small ~~near~~ the operating function part of sigmoid function will be linear & the model will collapse.

When the starting values are zero, it produce zero derivative & make perfect symmetry & the algorithm do not move further.

When the weights are large it will produce poor solution.

When the starting values are near zero it will be linear & when the weight increases it will become non-linear. For standarized i/p's, the starting values can be in the range $[+0.7, -0.7]$.

4. $L_p(\beta, \beta_0, \alpha_i) = \frac{1}{2}\|\beta\|^2 - \sum_{i=1}^{N} \alpha_i [y_i(x_i^T\beta + \beta_0) - 1]$

$$L_p(\beta, \beta_0, \alpha_i) = \frac{1}{2}\|\beta\|^2 - \sum_{i=1}^{N} \alpha_i y_i x_i^T \beta - \sum_{i=1}^{N} \alpha_i y_i \beta_0 + \sum_{i=1}^{N} \alpha_i$$

Differentiating L_p with respect to β & equating to 0.

$(x_i^T\beta = x_i \beta^T)$

$$\frac{\partial L_p}{\partial \beta} = \frac{1}{2} \times 2\beta - \sum_{i=1}^{N} \alpha_i y_i x_i = 0$$

$(\alpha > 0, y = 0$
$\alpha = 0, y > 0)$

$$\beta = \sum_{i=1}^{N} \alpha_i y_i x_i^T$$

Differentiating L_p with respect to β_0

$$\frac{\partial L_p}{\partial \beta_0} = \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\therefore \beta = \sum_{i=1}^{N} \alpha_i y_i x_i^T, \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

Substituting β in L_P forms the Wolfe-Dual

$$L_D(\beta, \beta_0, \alpha_i) = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k - \sum_{i=1}^{N}\sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k + \sum_{i=1}^{N} \alpha_i$$

$$= \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k$$

Now L_D can be represented in terms of α,

$$L_D(\alpha_i) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{N} \alpha_i \alpha_k y_i y_k x_i^T x_k$$

5. Steps followed for prediction using SVM.
   1. Transform data into the format of SVM package
   2. Conduct simple scaling on the data.
   3. Consider the RBF kernel, $K(x, x') = \exp(-\gamma ||(x - x')||^2)$
   4. Use cross-validation to find the best parameters of C & γ.
   5. Use the best parameters C & γ to train the whole training set.
   6. Test

6. $k(x, x') = (1 + x, x')^2 = (1 + x^T x')^2$

$$x^T = [x_1 \quad x_2], \quad x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$$

$$x^T x' = x_1 x_1' + x_2 x_2'$$

$$k(x, x') = (1 + x_1 x_1' + x_2 x_2')^2$$

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$$

where $a = 1, b = x_1 x_1', c = x_2 x_2'$

$$K(x, x') = (1 + 2x_1 x_1' + 2x_2 x_2' + (x_1 x_1')^2 + 2x_1 x_1' x_2 x_2' + (x_2 x_2')^2)$$

$$= [1 \quad \sqrt{2}x_1 \quad \sqrt{2}x_2 \quad x_1^2 \quad \sqrt{2}x_1 x_2 \quad x_2^2] \begin{bmatrix} 1 \\ \sqrt{2}x_1' \\ \sqrt{2}x_2' \\ x_1'^2 \\ \sqrt{2}x_1' x_2' \\ x_2'^2 \end{bmatrix}$$

$$= h(x^T)\, h(x')$$

The kernel function can be expressed as an inner product in a feature space mapping from 2 dimension to 6 dimension is by the mapping of input vector function $h(x)$

The components of $h(x)$ are $h(x_1) = 1$, $h(x_2) = \sqrt{2}x_1$, $h(x_3) = \sqrt{2}x_2$, $h(x_4) = x_1^2$, $h(x_5) = \sqrt{2}x_1x_2$ and $h(x_6) = x_2^2$