EE-559

HOMEWORK-7

1. To prove : for SVR,

$$w_0 = y_j - \varepsilon - \sum_{i \in S} (\lambda_i^+ - \lambda_i^-) K(x_i, x_j)$$

Proof:

By taking a derivative from Primal Lagrangian,

$\frac{\partial L}{\partial w} = 0$ we have, ref: slide 206.

$$w = \sum_{i=1}^{N} (\lambda_i^+ - \lambda_i^-) \varphi(x_i) = \sum_{i \in S} (\lambda_i^+ - \lambda_i^-) \varphi(x_i) \quad \text{—(i)}$$

Support vector condition:

for support vectors at the upper edge of ε-tube:

$$y_i - \hat{y}_i = \xi_i^+ + \varepsilon$$

since slack variables, on the edge $= 0$, we have in this case:

$$y_i - \hat{y}_i = \varepsilon$$

where $\hat{y}_i = w^T \varphi(x_i) + w_0$

Since question has y in terms of j, we can write.

$$\hat{y}_j = w^T \varphi(x_j) + w_0$$

$$y_j - \hat{y}_j = \varepsilon$$

$$\Rightarrow y_j - [w^T \varphi(x_j) + w_0] = \varepsilon$$

replacing w^T with (i)

$$y_j - \Big(\sum_{i \in S}^{N} (\lambda_i^+ - \lambda_i^-) \varphi(x_i)^T \varphi(x_j) + w_0\Big] = \varepsilon$$

applying kernel trick we get $\varphi(x_i)^T \varphi(x_j) \to K(x_i, x_j)$

$$y_j - \Big(\sum_{i \in S}^{N} (\lambda_i^+ - \lambda_i^-) K(x_i, x_j)\Big) + w_0 = \varepsilon$$

$$\Rightarrow w_0 = y_j - \varepsilon - \sum_{i \in S} (\lambda_i^+ - \lambda_i^-) K(x_i, x_j)$$

(b) Given: decision rule:

$$f \rightarrow Sign = \begin{cases} -1 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

Step 1:

$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad y_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

$$w_1(1) = w_1(0) + 0.5 \underbrace{e_1}_{1-y_1} x_1$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5 \left[1 - f(w_1^T(0)x_1)\right] x_1$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5 \left[1 - f(0)\right] x_1$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5 \left[1 - 1\right] x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Perceptron #2:

$$e_2 = -1 - f(\underbrace{w_2^T(0)x_2}_{0})$$

$$= -1 - f(0) = -1 - 1 = -2$$

$$\therefore w_2(1) = w_2(0) + 0.5 e_2 x_1$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5(-2) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

Perceptron 3

$$w_3(1) = w_3(0) + 0.5 e_3 x_1$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5 \left[-1 - f(0)\right] \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5 \left[-2\right] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\Rightarrow W(1) = \left[w_1(1) \quad w_2(1) \quad w_3(1)\right]$$

$$= \begin{bmatrix} 0 & -1 & -1 \\ 0 & -1 & -1 \end{bmatrix}$$

Data point 2:

$$x = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad y_2 = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$$

Perceptron 1:

$$w_1(2) = w_1(1) + 0.5\, e\, x$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5\left[-1 - f(w_1(1)^T x_2)\right]\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5\left[-1 - f\left([0 \;\; 0]\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)\right]\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5\,[-1 - 1]\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5(-2)\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\Rightarrow w_1(2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Perceptron 2:

$$w_2(2) = w_2(1) + 0.5\, e_2\, x_2$$

$$e = 1 - f(w_2(1)^T x_2) = 1 - f\left[[-1 \;\; -1]\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right] = 1 - f(0) = 1 - 1 = \underline{\underline{0}}$$

$$\therefore w_2(2) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + 0.5\,(0)\cdot\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$w_2(2) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

Perceptron 3

$$w_3(2) = w_3(1) + 0.5\, e\, x_2$$

$$e = -1 - f(w_3(1)^T x_2) = -1 - f\left[[-1 \;\; -1]\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right] = -1 - f(0)$$

$$= -1 - 1 = -2$$

$$\therefore w_3(2) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + 0.5\,(-2)\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\Rightarrow W(2) = \begin{bmatrix} -1 & -1 & -2 \\ 1 & -1 & 0 \end{bmatrix}$$

3: $x = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $y_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$

$$w_1(3) = w_1(2)^{\cancel{T}} + 0.5\, e x_3$$

$$e = -1 - f\left(\left[w_1(2)^T x_3\right]\right) = -1 - f\Big(\underbrace{\begin{bmatrix} -1 & +1 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}}_{1 + (-1) = 0}\Big)$$

$$e = -1 - f(0) = -1 - 1 = -2$$

$$\therefore w_1(3) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0.5(-2)\begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

$$w_2(3) = w_2(2) + 0.5\, e x_3$$

$$e = -1 - f\left(\left[w_2(2)^T x\right]\right) = 1 - f\Big(\underbrace{\begin{bmatrix} -1 & -1 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}}_{1 + 1 = 2}\Big)$$

$$e = 1 - f(2) \Rightarrow e = -(1+1) = -2$$

$$w_2(3) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + 0.5(-2)\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$w_3(3) = w_3(2) + 0.5\, x_3 e$$

$$e = 1 - f\left(\left[w_3(2)^T x\right]\right) = 1 - f\Big(\underbrace{\begin{bmatrix} -2 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ -1 \end{bmatrix}}_{2 + 0}\Big)$$

$$e = 1 - f(2) = 1 - 1 = 0$$

$$\therefore w_3(3) = \begin{bmatrix} -2 \\ 0 \end{bmatrix} + 0.5(0)\begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

$$\Rightarrow w(3) = \begin{bmatrix} 0 & 0 & -2 \\ 2 & 0 & 0 \end{bmatrix}$$

Second epoch: $(\underset{x_1}{w_4}, \underset{x_2}{w_5}, \underset{x_3}{w_6})$

for $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $y_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

$w_1(4) = w_1(3) + 0.5\, e\, x_1$

$e = 1 - f([w_1(3)^T x_1]) = 1 - f(\underbrace{[0 \;\; 2]\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{0+2=2})$

$e = 1 - f(2) = 1 - 1 = 0$

$\therefore w_1(4) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.5(0)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

$w_2(4) = w_2(3) + 0.5\, x\, e$

$e = -1 - f([w_2(3)^T x_1]) = 1 - f(\underbrace{[0 \;\; 0]\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{0}) = 1 - f(0) = -1 - 1 = -2$

$\Rightarrow w_2(4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0.5(-2)\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\Rightarrow w_2(4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$w_3(4) = w_3(3) + 0.5\, x\, (e)$

$e = -1 - f([w_3(3)^T x_1]) = -1 - f(\underbrace{[-2 \;\; 0]\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{-2+0}) = -1 - f(-2)$
$= -1 - (-1) = 0$

$\Rightarrow w_3(4) = \begin{bmatrix} -2 \\ 0 \end{bmatrix} + 0.5(0)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$

$\Rightarrow W_4 = \begin{bmatrix} 0 & -1 & -2 \\ 2 & -1 & 0 \end{bmatrix}$

Step 5 using $x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $y_2 = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}$

$w_1(5) = w_1(4) + 0.5\, e\, x_1$

$e = -1 - f([w_1(4)^T x_1]) = -1 - f(\underbrace{[0 \;\; 2]\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{0+(-2)=-2}) = -1 - \underbrace{f(-2)}_{-1} = -1 - (-1) = 0$

$w_1(5) = \begin{bmatrix} 0 \\ 2 \end{bmatrix} + 0.5(0)\begin{bmatrix} +1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

$w_2(5) = w_2(4) + 0.5\, x_2 e$

$e = 1 - f([w_2(4)^T x_2]) = 1 - f(\underbrace{[-1\ \ -1]\begin{bmatrix}1\\-1\end{bmatrix}}_{-1+(1)=0}) = 1 - f(0) = 1-1 = 0$

$w_2(5) = \begin{bmatrix}-1\\-1\end{bmatrix} + 0.5(0)\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}-1\\-1\end{bmatrix}$

$w_3(5) = w_3(4) + 0.5\, x_2 e$

$e = -1 - f([w_3(4)^T x_2]) = -1 - f(\underbrace{[-2\ \ 0]\begin{bmatrix}1\\-1\end{bmatrix}}_{-2+0=-2}) = -1 - f(-2) = -1-(-1) = 0$

$w_3(5) = \begin{bmatrix}-2\\0\end{bmatrix} + 0.5(0)\begin{bmatrix}+1\\-1\end{bmatrix} = \begin{bmatrix}-2\\0\end{bmatrix}$

$\Rightarrow w(5) = \begin{bmatrix}0 & -1 & -2\\2 & -1 & 0\end{bmatrix}$

<u>Step 6</u>: Using $x_3 = \begin{bmatrix}-1\\-1\end{bmatrix}$ $y_3 = \begin{bmatrix}-1\\-1\\1\end{bmatrix}$

$w_1(6) = w_1(5) + 0.5\,(e)\, x_3$

$e = -1 - f([w_1(5)^T x_3]) = -1 - f(\underbrace{[0\ \ 2]\begin{bmatrix}-1\\-1\end{bmatrix}}_{0\ -2}) = -1 - f(-2) = -1+1 = 0$

$w_1(6) = \begin{bmatrix}0\\2\end{bmatrix} + 0.5(0)\begin{bmatrix}-1\\-1\end{bmatrix} = \begin{bmatrix}0\\2\end{bmatrix}$

$w_2(6) = w_2(5) + 0.5\,(e)\, x_3$

$e = -1 - f([w_2(5)^T x_3]) = -1 - f(\underbrace{[-1\ \ -1]\begin{bmatrix}-1\\-1\end{bmatrix}}_{1+1=2}) = 1 - f(2) = -1-1 = -2$

$w_2(6) = \begin{bmatrix}-1\\-1\end{bmatrix} + 0.5(-2)\begin{bmatrix}-1\\-1\end{bmatrix} = \begin{bmatrix}-1\\-1\end{bmatrix} + \begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

$w_3(6) = w_3(5) + 0.5\,(e)\, x_3$

$e = +1 - f([-2\ \ 0]\begin{bmatrix}-1\\-1\end{bmatrix}) = 1 - f(2) = 1-1 = 0$

$w_3(6) = \begin{bmatrix}-2\\0\end{bmatrix} + 0.5(0)\begin{bmatrix}-1\\-1\end{bmatrix} = \begin{bmatrix}-2\\0\end{bmatrix}$

$\Rightarrow w(6) = \begin{bmatrix}0 & 0 & -2\\2 & 0 & 0\end{bmatrix}$ $\rightarrow$ Similar to $w(3)$
(=)

Convergence:

Using $W(6) = \begin{bmatrix} 0 & 0 & -2 \\ 2 & 0 & 0 \end{bmatrix}$

for $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\quad y_1 = \begin{array}{ccc} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} & \begin{array}{c} w_1(6) \\ w_2(6) \\ w_3(6) \end{array} & \begin{array}{c} y_{11} \\ y_{12} \\ y_{13} \end{array} \end{array}$

$w_1(6)^T x_1 = [0 \;\; 2]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 \quad Sign(2) = 1 = y_{11}$ ✓

$w_2(6)^T x_1 = [0 \;\; 0]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \quad Sign(0) = 1 \neq y_{12}$ ✗

$w_3(6)^T x_1 = [-2 \;\; 0]\begin{bmatrix} 1 \\ 1 \end{bmatrix} = -2 \quad Sign(-2) = -1 = y_{13}$ ✓

---

for $x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $\quad y_2 = \begin{array}{cc} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} & \begin{array}{c} y_{21} \\ y_{22} \\ y_{23} \end{array} \end{array}$

$w_1(6)^T x_2 = [0 \;\; 2]\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 2 \Rightarrow Sign(2) = 1 = y_{21}$ ✓

$w_2(6)^T x_2 = [0 \;\; 0]\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0 \Rightarrow Sign(0) = 1 = y_{22}$ ✓

$w_3(6)^T x_2 = [-2 \;\; 0]\begin{bmatrix} +1 \\ -1 \end{bmatrix} = -2 \Rightarrow Sign(-2) = -1 = y_{23}$ ✓

---

for $x = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $\quad y_3 = \begin{array}{cc} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} & \begin{array}{c} y_{31} \\ y_{32} \\ y_{33} \end{array} \end{array}$

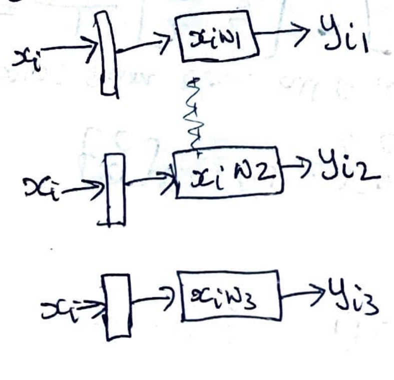
$w_1(6)^T x_3 = [0 \;\; 2]\begin{bmatrix} -1 \\ -1 \end{bmatrix} = -2 \Rightarrow Sign(-2) = -1 = y_{31}$ ✓

$w_2(6)^T x_3 = [0 \;\; 0]\begin{bmatrix} -1 \\ -1 \end{bmatrix} = 0 \Rightarrow Sign(0) = 1 \neq y_{32}$ ✗

$w_3(6)^T x_3 = [-2 \;\; 0]\begin{bmatrix} -1 \\ -1 \end{bmatrix} = 2 \Rightarrow Sign(2) = 1 = y_{33}$ ✓

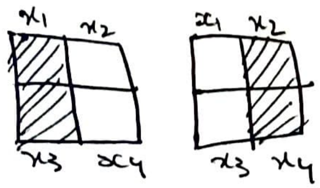
7/9 correctly classified ⇒ there is still error and weights have not converged.

architecture:

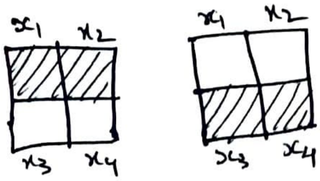


performed over x_i's.

3)



} class 1



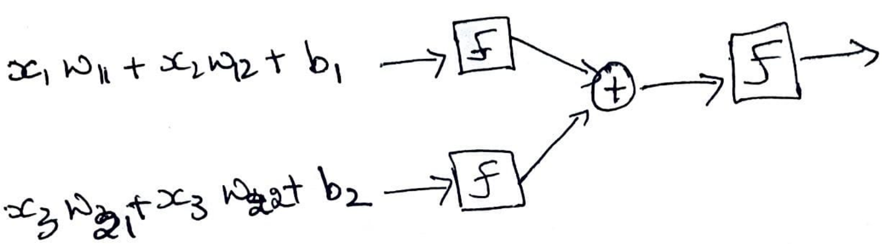
} class 2

First layer : two neurons : AND operations (O/P=1 when both i/ps $\to$ 1)

Second : One neuron $\to$ OR C

Pattern 1: $\to$ {x_1 x_3 or x_2 x_4} (class 1).

Class 2: $\Rightarrow$ {x_1 x_2 or x_3 x_4}



class 1:

$$\begin{array}{cc} x_1 & x_2 \end{array} \qquad \begin{array}{cc} x_1 & x_2 \end{array}$$
$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$
$$\begin{array}{cc} x_3 & x_4 \end{array} \qquad \begin{array}{cc} x_3 & x_4 \end{array}$$

class 2:

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

Inputs:

class 1:

$X = [1\,0\,1\,0]^T$ (or)

$[0\,1\,0\,1]^T$

$X = [1\,1\,0\,0]^T$ or

$[0\,0\,1\,1]^T$

output:

$0 \to$ class 1

$1 \Rightarrow$ class 2.

First hidden layer: (AND) if $w_{11} = w_{12} = w_{21} = w_{22} = 1$ & $b_1 = b_2 = 0.5$

$$z_1 = step(x_1 + x_2 + 0.5)$$
$$z_2 = step(x_3 + x_4 + 0.5)$$

Eg: class = $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow [1\;1\;0\;0]$

$x_1 = x_2 = 1 \qquad x_3 = x_4 = 0$

$\therefore Z_1 = \text{step}(1+1+0.5) = 1 \xrightarrow{OR}$

$Z_2 = \text{step}(0+0+0.5) = 1$

Let's choose: $w_{31} = w_{32} = 1$, bias $= -2$

$$Z_3 = \text{step}(w_{31} Z_1 + w_{32} Z_2 + b)$$

$$= \text{step}(1+1-2) = 0 \text{ incorrect } \times.$$

following above similar iterations,
optimal weights got are

$$\left.\begin{array}{l} w_{11} = w_{12} = w_{21} = w_{22} = 1 \\ b = -1.5 \end{array}\right\} \text{~~OR~~ AND layer}$$

$$\left.\begin{array}{l} w_{31} = w_{32} = 1 \\ b = -0.5 \end{array}\right\} \text{OR layer.}$$

Architecture: