1001553524 Varsha Rani chawan

HW2

(1a) The given problem corresponds to linear binary classifier

① optimization function & constraints for linear maximum margin optimization

→ the hypothesis space that seperates the data point is

$$h_\theta(x) = \theta^T x$$

$$h_{wb}(x) = w^T x + b$$

→ two kinds of margins are formulated as

① functional margin

$$\hat{\gamma}^i = y^i (w^T x^{(i)} + b)$$

→ it has a scaling problem if used with logistic regression. It is not a good measure of confidence.

② geometric margin;

$$\gamma^i = y^i \left(\frac{w^T}{||w||} \cdot x^i + \frac{b}{||w||} \right)$$

when $||w|| = 1$, functional margin = geometric margin

~~1b~~ (ic) when $\|w\| = 1$

$$\gamma = \hat{\gamma} / \|w\|$$

→ as the samples given are linearly seperable, the optimization of SVM is to find a hyperplane that seperates the ~~the~~ positive and the negative points with the maximum gap. ∴ the optimization function is geometric margin.

$$\Rightarrow \max_{\gamma, w, b} \gamma / \|w\|$$

$$\text{s.t } y^i (w^T x^i + b) \geq \gamma, \quad i = 1 \dots n$$

$$\|w\| = 1$$

↳ constrained optimization problem

~~→ scale the~~

for non-convex constraints

$$\max_{\gamma, w, b} \frac{\hat{\gamma}}{\|w\|}$$

$$\text{s.t } y^i (w^T x^i + b) \geq \hat{\gamma}, \quad i = 1 \dots n$$

scale the margin

$$\Rightarrow \min_{\gamma, w, b} \frac{1}{2} \|w\|^2$$

$$\text{s.t } y^i (w^T x^i + b) \geq 1 \quad i = 1 \dots n$$

(ii) Lagrangian :-

Generalized Lagrangian is given by considering both equality & inequality constraints as

$$\mathcal{L}(w, \alpha, \beta) = f(w) + \sum_{i=1}^{K} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$$

→ By considering the Lagrange Duality:

Lagrangian is given as:

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y^i (w^T x^i + b) - 1 \right]$$

→ Lagrangian dual problem is given by

$$\max_\alpha \quad W(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} y^i y^j \alpha_i \alpha_j \langle x^i, x^j \rangle$$

$$\text{s.t} \quad \alpha_i \geq 0 \,, \quad i = 1 \ldots m$$

$$\sum_{i=1}^{m} \alpha_i y^{(i)} = 0.$$

~~Onle~~ ~~we~~ ~~solu~~ solve α^* by SMO

then compute

$$w = \sum_{i=1}^{m} \alpha_i y^i x^i$$

$$b^* = -\frac{\max_{i:y^{(i)}=-1} w^T x^i + \min_{i:y^{(i)}=1} w^T x^i}{2}$$

1b SMO algorithm:

Pseudo code:

Below is the pseudo code for the calculations done, to perform 4 iterations of SMO

$X = ([[1,2],[2,3],[2,1],[3,4],[1,3],[4,4]])$

$Y = [-1,1,-1,1,-1,1]$

- initialize $\alpha_i = [0\ 0\ 0\ 0\ 0\ 0]$ $b=0$, $C=1$

repeat for 4 iterations

- for (count = 4)
  - $\alpha_i^{old} = \alpha_i$
  - for ($j = 0$ to 5) [as x has 6 values]
    - select random i between [0,5] [$i \neq j$]
      - $\hat{\alpha}_i, \hat{\alpha}_j = \alpha_i, \alpha_j$ [Save old α]
      - based on i & j values get the corresponding x_i, x_j, y_i, y_j
      - Calculate L and H by the below formula.

if $(y_i \neq y_j) \Leftarrow$ $L = \max(0, \hat{\alpha}_j - \hat{\alpha}_i)$, $H = \min(C, C + \hat{\alpha}_j - \hat{\alpha}_i)$

If $y_i = y_j$

$L = \max(0, \hat{\alpha}_i + \hat{\alpha}_j - C)$, $H = \min(C, \hat{\alpha}_i + \hat{\alpha}_j)$

- Calculate weights & bias.

$$w = X^T \cdot (\alpha \times Y) \Rightarrow \left(\sum_{i=1}^{m} x_i^T \alpha_i y_i \right)$$

$$b = \underline{Y - (w^T X^T)}$$

Then find the average of 'b' to get single b value

- Calculate $E_K = f(x_K) - y_K$

$$E_i = \left[(w^T x_i^T) + b \right] - y_i$$

$$E_j = \left[(w^T x_j^T) + b \right] - y_j$$

- Calculate α_j

$$\alpha_j = \hat{\alpha}_j - \frac{y^j \left[E_i - E_j \right]}{\eta}$$

for this calculate η

- $\eta = 2(x_i, x_j) - (x_i x_i) - (x^j x^j)$

clip α_j to lie with in range $[L, H]$

$$\alpha_j := \max(\alpha_j, L)$$

$$\alpha_j = \min(\alpha_j, H)$$

- Calculate α_i

$$\alpha_i = \hat{\alpha}_i + y_i y_j (\hat{\alpha}_j - \alpha_j)$$

repeat this for all $i = 1$ to 6 ((ie) 1 to m)
after every iteration copy the α values

after end of iteration 4
take α^* & find the weight & bias for ~~using~~ the final value.

Calculations :-

I have mentioned the steps & formulae used in the below calculations in the ~~PSP~~ pseudo code, so I am directly writing the calculation values.

$\alpha = [0\ 0\ 0\ 0\ 0\ 0]$, $C = 1$

Iteration 1

- $j = 0$, randomly took $i = 1$

$\Rightarrow x_i = [2, 3]$, $x_j = [1, 2]$ $y_i = 1$, $y_j = -1$

$\hat{\alpha}_i , \hat{\alpha}_j = 0, 0$

$L, H = [\max[0, 0], \min[1, 1]]$

$\because y_i \neq y_j$

$\therefore L, H = [0, 1]$

$w = [X^T] \cdot [\alpha\, y]$

$$= \begin{bmatrix} 1 & 2 & 2 & 3 & 1 & 4 \\ 2 & 3 & 1 & 4 & 3 & 4 \end{bmatrix}_{2\times 6} \begin{array}{c} \alpha \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{array} \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

$\Rightarrow$ $w = [0, 0]$

bias $= b = y - [w^T x^T]$

$$= \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 2 & 1 \\ 3 & 4 \\ 1 & 3 \\ 4 & 4 \end{bmatrix}^T [0, 0]$$

$$= [-1 \; 1 \; -1 \; 1 \; -1 \; 1]$$

mean $b = \frac{0}{6} = 0$

$\therefore b = 0$

- Calculating E values: $x_i = [2, 3]$, $w = [0, 0]$, $b = 0$

$$E_i = \left[[w^T x_i^T] + b \right] - y_i \qquad y_i = +1$$

$$= 0 - (+1) = -1$$

$$E_j = 0 - (-1) = 1 \qquad \because y_j = -1$$

$$\eta = (x_i x_i) + (x_j x_j) - 2(x_i x_j)$$

$$= [2,3][2,3] + [1,2][1,2] - 2[2,3][1,2]$$

$$= 13 + 5 - 2 \times 8 = 2$$

- Calculate α_j

$$\alpha_j = \hat{\alpha}_j + y_j * \frac{[E_i - E_j]}{\eta}$$

$$\alpha_j = 0 + (-1) \times \frac{[0 - 1]}{2}$$

$$\alpha_j = 1$$

Chek the conditions it lies with in [0,1]

$\therefore \alpha_i = \hat{\alpha_i} + y_i y_j [\hat{\alpha_j} - \alpha_j]$

$\alpha_i = 0 + [-1][0-1]$

$\alpha_i = 1$

now repeat above steps for

$j = 1$, choose $i = 2$ randomly

$x_i = [2,1]$, $x_j = [2,3]$, $y_j = -1$, $y_j = 1$

$\hat{\alpha_i} = 1$, $\hat{\alpha_j} = 1$

$L, H = [1,1]$ [I am writing direct values]

$w = [1,1]$

$b = [-4, -4, -4, -6, -5, -7]$

$b = -5$ / average b

$E_i = 0$, $E_j = -1$

$\therefore \eta = 4$

$\alpha_j = 1.25$ [clip the α_j to 1]

$\alpha_i = 1$

now again repeat the steps for $j = 2$, $j = 3$

$j = 4$, $j = 5$

then at the end of iteration 1

we get $\alpha = [1\,1\,0\,0\,0\,0]$

now reapeat same for iteration 4

end of iteration 2 : $\alpha_2 = [1, 1, 0, 0, 0, 0]$

iteration 3 : $\alpha_3 = [1, 1, 0, 0, 0, 0]$

iteration 4 : $\alpha_4 = [1, 1, 0, 0, 0, 0]$

now find final w & b with the final α Value

$w = [1, 1]$, $b = -5.0$ //

(*) If we do not consider any conditions to choose α_i & α_j as given in assignment then following the algorithm pseduo code Just skip the logic to clip α_j we get the below α Values after each iteration: (with out heuristic)

iteration 1
$\alpha_1 = [0, 0.5, 0.1, 0, 0.4, 0]$

$\alpha_2 = [0, 2.5, 0.1, 0, 2.4, 0]$

iteration 3 → $\alpha_3 = [1, 3.5, 0.1, 0, 2.4, 0]$

iteration 4 → $\alpha_4 = [1, 6, 0.6, 0, 4.4, 0]$

$\Rightarrow$ $b = -22.266$
$w = [7.4, 2.2]$
} at the end of iteration 4