MACHINE LEARNING

ASSIGNMENT 3 REPORT

AASTHA 2019224

Date________
Page No.______

2 A) No, there should not be a non-negativity constraint on ε for l2 norm soft margin SVM optimization. The ε_i term is squared and will always be positive. The presence/absence of this non-negativity constraint will not affect the optimum value.

If $\varepsilon_i < 0$, then $1 - \varepsilon_i \geq 1$

and hence $y^{(i)} (w^T x^{(i)} + b) \geq 1$

Therefore $y^{(i)} (w^T x^{(i)} + b) \geq 1 - 0$

$$y^{(i)} (w^T x^{(i)} + b) \geq 1 - \varepsilon_i \quad , \varepsilon_i = 0$$

hence the constraint is valid even for $\varepsilon_i = 0$ if $\varepsilon_i < 0$. The objective function $\min_{w,b,\varepsilon} \frac{1}{2} \|w\|^2 + \frac{c}{2} \sum_{i=1}^{m} \varepsilon_i^2$ will have a lower value. Hence this cannot be optimum solution for the optimization problem.

$\Rightarrow$ The non negativity constraint is not required.

Date
Page No.

B) The Lagrangian of the above optimization problem is:
(Primal form, $\alpha_i \geq 0$)

$$L_p(W, b, \varepsilon, \alpha) = \frac{1}{2} W^T W + \frac{C}{2} \sum_{i=1}^{m} \varepsilon_i^2$$

$$- \sum_{i=1}^{m} \alpha_i \left[y^{(i)} \left(W^T x^{(i)} + b \right) - 1 + \varepsilon_i \right]$$

for $i = 1, \ldots, m$

c) Conditions for optimality

Karush-Kuhn Tucker Condition must be satisfied.

$$\alpha_i \left(y_i (x_i \cdot W + b) - 1 + \varepsilon_i \right) = 0$$

$i = 1, \ldots m$

$$\frac{\partial L_p(W, b, \varepsilon, \alpha)}{\partial W} = 0 \quad -(1)$$

$$\frac{\partial L_p(W, b, \varepsilon, \alpha)}{\partial b} = 0 \quad -(2)$$

$$\frac{\partial L_p(W, b, \varepsilon, \alpha)}{\partial \varepsilon_i} = 0 \quad -(3)$$

Finding the derivatives

$$\frac{\partial L_p(W, b, \varepsilon, \alpha)}{\partial W} = W - \sum_{i=1}^{m} \alpha_i \beta_i x_i \quad -(4)$$

Date
Page No.

$$\frac{\partial L_p(W, b, \mathcal{E}, \alpha)}{\partial b} = \sum_{i=1}^{m} y_i \alpha_i \quad (5)$$

$$\frac{\partial L_p(W, b, \mathcal{E}, \alpha)}{\partial \mathcal{E}_i} = C\mathcal{E}_i - \alpha_i \quad (6)$$

from (1), (2), (3) & (4), (5), (6) we get

$$W = \sum_{i=1}^{m} \alpha_i \beta_i x_i$$

$$\sum_{i=1}^{m} y_i \alpha_i = 0$$

$$C\mathcal{E}_i = \alpha_i$$

substituting these in primal form. we get the dual of the above optimization problem:

maximize

$$\theta(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$- \frac{1}{2} \sum_{i=1}^{m} \frac{\alpha_i^2}{C}$$

subject to $\sum_{i=1}^{m} \alpha_i y^{(i)} = 0$, $\alpha_i \geq 0$ for

$i = 1, \ldots, m$

Date
Page No.

3 A) Using the hint provided

Let $\alpha_i = 1$ for all i and $b=0$, we notice that for $y \in \{-1, 1\}$ prediction on $x(i)$ will be correct if $|f(x^{(i)}) - y^{(i)}| < 1$
We need to find a value for τ that satisfies the inequality for all i.

$$|f(x^{(i)}) - y^{(i)}| = \left| \sum_{i=1}^{m} \alpha_i y^{(i)} K(x^{(i)}, x) + b - y^{(i)} \right|$$

$\alpha_i = 1$ and $b = 0$, hence

$$= \left| \sum_{i=1}^{m} y^{(i)} K(x^{(i)}, x) - y^{(i)} \right|$$

$$= \left| \sum_{j=1}^{m} y^{(j)} \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2} \right) - y^{(i)} \right|$$

$$= \left| \sum_{j \neq i}^{m} y^{(j)} \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2} \right) \right.$$

$$\left. + \sum_{j=i} y^{(j)} \exp(0) - y^{(i)} \right|$$

$$= \left| \sum_{j \neq i} y^{(j)} \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2} \right) + y^{(i)} - y^{(i)} \right|$$

$$= \left| \sum_{j \neq i} y^{(j)} \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2} \right) \right.$$

Date__________
Page No.________

Taking mod inside the summation

$$\left|f(x^{(i)}) - y^{(i)}\right| \le \sum_{j \neq i} |y^{(j)}| \left| \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2}\right) \right|$$

(triangle inequality)

$|y^{(i)}| = 1$

$$= \sum_{j \neq i} \exp\left(\frac{-\|x^{(j)} - x^{(i)}\|^2}{\tau^2}\right)$$

It is given that $\|x(j) - x(i)\| \ge \epsilon$ for any $i \neq j$

$$\le \sum_{j \neq i} \exp\left(-\frac{\epsilon^2}{\tau^2}\right)$$

$$= (m-1) \exp\left(-\frac{\epsilon^2}{\tau^2}\right)$$

We need to find a value of τ for which

$$\left|f(x^{(i)}) - y^{(i)}\right| < 1$$

$$\Rightarrow \quad (m-1) \exp\left(-\frac{\epsilon^2}{\tau^2}\right) < 1$$

$$\log\left((m-1)\exp\left(\frac{-\varrho^2}{\tau^2}\right)\right) < \log 1$$

$$\log(m-1) - \frac{\varrho^2}{\tau^2} < 0$$

$$\log(m-1) < \frac{\varrho^2}{\tau^2}$$

$$\tau^2 < \frac{\varrho^2}{\log(m-1)}$$

$\tau^2 = \dfrac{\varrho^2}{2\log(m-1)}$ satisfies this inequality.

Date
Page No.

3 B) Yes, we will obtain zero training error on resulting classifier.

If we find a solution for the SVM then it has zero training error.

The optimum values should satisfy the constraints and let $b=0$

$$y^{(i)} (W^T x^{(i)} + b)$$

$$= y^{(i)} f(x^{(i)})$$

This will be greater than 0, since the terms being multiplied have the same sign.

Choosing big values of α_i, $y^{(i)} f(x^{(i)})$ can be greater than 1.
Hence

$$y^{(i)} (W^T x^{(i)} + b) > 1$$

Therefore, we can obtain an optimum solution and zero training error.